DOCUMENT RESUME

ED 142 457 SO 010 103

AUTHOR	Tikunoff, William J., Ed.; Ward, Beatrice A., Ed.
TITLE	Anthropology and Education Quarterly, Vol. VIII, No. 2, May 1977. Special Issue: Exploring Qualitative/Quantitative Research Methodologies in Education.
INSTITUTION	Council on Anthropology and Education, Washington, D.C.
PUB DATE	May 77
NOTE	131p.
AVAILABLE FROM	Council on Anthropology and Education, 1703 New Hampshire Avenue, N.W., Washington, D.C. 20009 ($2.00 paper cover)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Classroom Environment; Conference Reports; Data Collection; Demonstration Projects; *Educational Anthropology; *Educational Research; Effective Teaching; Language Development; Race Relations; Research Design; *Research Methodology; Research Problems; Scientific Methodology
IDENTIFIERS	Ethnography

ABSTRACT

This special journal issue contains the proceedings of the Workshop Exploring Qualitative/Quantitative Research Methodologies in Education, held July 21-23, 1976, in Monterey, California. An introductory overview comments on the epistemological nature of the quantitative and the qualitative approaches to educational research and compares several dominant motifs and patterns that serve to clarify the alternative emphases inherent in each approach. The papers that follow address five educational problems: (1) determining next steps in qualitative data collection; (2) assessing language development--written and/or oral; (3) examining reasons for doing demonstration projects; (4) identifying effective teaching; and (5) assessing race relations in the classroom. There are three papers per problem--the first from a researcher working in a qualitative mode, the second from an expert working in the quantitative mode, and the third a critique of the first two papers by a person who represents expertise in the problem area identified either as a researcher or a practitioner. (Author/RM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED142457

JUN 16 1977

ANTHROPOLOGY & EDUCATION QUARTERLY

VOLUME VIII, NUMBER 2

MAY 1977

SPECIAL ISSUE: EXPLORING QUALITATIVE/QUANTITATIVE RESEARCH METHODOLOGIES IN EDUCATION*

ISSUE EDITORS: WILLIAM J. TIKUNOFF and BEATRICE A. WARD
Far West Laboratory for Educational Research and Development

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

CAE; Chilcott

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

37 / **FOREWORD**
Robert B. Textor

40 / **INTRODUCTION**
William J. Tikunoff and *Beatrice A. Ward*

OVERVIEW

42 / **On the Relations Among Educational Research Paradigms: From Disdain to Detente**
Ray C. Rist

NEXT STEPS IN QUALITATIVE DATA COLLECTION

50 **The Collection and Analysis of Ethnographic Data in Educational Research**
Stephen E. Fienberg

58 / **Some Approaches to Inquiry in School-Community Ethnography**
Frederick Erickson

70 / **Critiques** by *George W. Fairgrieve* and *Elliot W. Eisner*

ASSESSING LANGUAGE DEVELOPMENT—WRITTEN AND/OR ORAL

73 / **Quantitative Language Data: A Case for and Some Warnings Against**
Roger W. Shuy

83 / **Language Assessment: Where, What and How**
Courtney B. Cazden with *James T. Bond*, *Ann S. Epstein*, *Robert D. Matz*, and *Sandra J. Savignon*

91 / **Critiques** by *Dell H. Hymes* and *Blandina Cardenas*

WHY DO DEMONSTRATIONS?

95 / **Why Do Demonstration Projects?**
John E. Dawson

106 / **Ethnographic Case Studies in Federally Funded Multi-Disciplinary Policy Research: Some Design and Implementation Issues**
Robert E. Herriott

115 / **Critiques** by *Walter J. Symons* and *Gene E. Hall*

HOW TO IDENTIFY EFFECTIVE TEACHING

119 / **How to Identify Effective Teaching**
William W. Cooley, *Gaea Leinhardt*, and *Janet McGrail*

127 / **Effective Teaching: A Qualitative Inquiry in Aesthetic Education**
Louis M. Smith

140 / **Critique** by *Gary A. Griffin*

ASSESSING RACE RELATIONS IN THE CLASSROOM

142 / **Racial Tension in High Schools: Pushing the Survey Method Closer to Reality**
Robert L. Crain

152 / **Race and the "We They Dichotomy" in Culture**
Eleanor B. Leacock

159 / **Critiques** by *Gladys G. Handy* and *Robert E. Wentz*

162 / **ATTENDEES**

SO010103

---

*Derived from a Workshop in Monterey, California, July 1976, sponsored by the Far West Laboratory for Educational Research and Development, in cooperation with the National Institute of Education and the Council on Anthropology and Education. The Workshop Exploring Qualitative/Quantitative Research Methodologies in Education was supported by the National Institute of Education, Department of Health, Education and Welfare, Contract No. NE-C-00-3-0108. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by the National Institute of Education should be inferred.

ANTHROPOLOGY AND EDUCATION QUARTERLY

(formerly *Council on Anthropology and Education Quarterly*,
incorporating *Council on Anthropology and Education Newsletter*)

ANNOUNCEMENT OF NEW EDITOR

The new editor of the *Anthropology and Education Quarterly* is Charles Harrington of Columbia University. Manuscripts (see Information for Authors) and editorial correspondence should be sent to him directly.

Charles Harrington, Editor
Anthropology and Education Quarterly
Box 19
Teachers College, Columbia University
New York, NY 10027

Subscription to *AEQ* is by membership in the Council on Anthropology and Education (annual dues for individual members $7.50) or by library subscription ($10.00 per annum). Single copies, when available, $2.00 each.

INFORMATION FOR AUTHORS

Manuscripts *(in triplicate)* should be addressed to the Editor of the *Anthropology and Education Quarterly*, Box 19, Teachers College, Columbia University, New York, New York 10027. All papers appropriate for the AEQ are sent anonymously to readers. Only the title should appear on the manuscript itself; all identifying material (name and affiliation) should be restricted to a cover page which will be removed when the paper is sent for review. All identifying references and footnotes should appear on a separate removeable page.

* * *

An abstract of 50-75 words must accompany each article, followed by a list of not more than five words or phrases under which the article can be indexed.

* * *

In all other aspects as to style and reference, manuscripts must follow the style guide of the *American Anthropologist*, printed in the September 1976 issue (Vol. 78, No. 3, pp. 747-752).

THE COUNCIL ON ANTHROPOLOGY AND EDUCATION

The Council on Anthropology and Education (CAE) is a professional association of anthropologists and educational researchers concerned with the application of anthropology to research and development in education. Organized in 1968 within the American Anthropological Association, but separately incorporated, the Council regularly meets at the annual meetings of the Association.

Address all membership and subscription communications to:

Anthropology and Education Quarterly
Council on Anthropology and Education
1703 New Hampshire Avenue, N.W.
Washington, DC 20009

Copyright © 1977 by the Council on Anthropology and Education. All rights reserved. No part of this publication may be reprinted in any form or by any means without prior written permission from the publisher.
Published quarterly—February, May, August, November—by the Council on Anthropology and Education, 1703 New Hampshire Ave., N.W., Washington, DC 20009. Printed in U.S.A. Second class postage paid at Washington, DC.

FOREWORD

Robert B. Textor
Stanford University

It is a pleasure to write this Foreword to the proceedings of the Workshop Exploring Qualitative/Quantitative Research Methodologies in Education, held July 21-23, 1976, in Monterey, California. I write from the vantage point of having been the representative of the Council on Anthropology and Education (CAE) who served on the Planning Committee for the Workshop. In this role, I was asked to represent the qualitative, ethnographic perspective in the design of the Workshop.

Throughout, the driving force behind the Workshop came from the Far West Laboratory for Educational Research and Development (FWL) in the persons of two educational researchers, William J. Tikunoff and Beatrice A. Ward. They were buoyed by their own recent successful experience in the use of one kind of ethnography, in combination with quantitative methods, in identifying teacher behaviors and classroom climates conducive to effective learning, and their enthusiasm was contagious. William and Betty were *primus inter pares* on the planning Committee, and the present publication bears the stamp of their concerns. Rounding out the Planning Committee was Dr. Virginia Koehler of the National Institute of Education (NIE), whose concern and tangible support made the entire venture possible.

Initial plans for the Workshop were made at a meeting at NIE in Washinton in November 1975. Besides the Planning Committee, others who participated from NIE included Ray C. Rist, himself an enthusiastic educational sociologist and ethnographer, John Schwille, and Andrew Porter; all three of these scholars agreed to serve as Advisors to the Planning Committee. John D. Herzog, who the following month succeeded me as President of the CAE, also agreed to serve as Advisor, and did so actively and creatively.

At the Washington meeting I suggested that the CAE might wish to co-sponsor the Workshop, and that the *Anthropology and Education Quarterly* would be a suitable outlet for publishing the proceedings. Those present agreed, as, later, did the Board of Directors of the CAE. The present issue of the *Quarterly* is the result: an issue more than twice as long as any previous one, and the first issue more than twice as long as any previous one, and the first issue ever to be underwritten by an outside organization—thanks to NIE and FWL. Moreover, the present issue is the first ever which, by advance planning, will be distributed to many *more* educational researchers, planners, and policy-makers *outside* the CAE than inside—again, thanks to NIE and FWL. In writing this Foreword, then, I am attempting to communicate not only with the regular CAE membership, but also with our special readership for this issue, for whom terms like "ethnography" might be somewhat unfamiliar.

The Planning Committee laid out a format calling for a paper by a qualitative and by a quantitative researcher on each of several broad areas of concern to American educational policy-makers, each paper then to be discussed by two discussants. In suggesting names of qualitative researchers who would be appropriate to invite, I consulted closely with John Herzog and Frederick Erickson, the other two members of the CAE Executive Committee, and with every past president of the CAE whom I could reach. A long list of nominees was thus assembled. Then came the task of fitting nominees to the constraints of each subject matter area, and matching the qualitative nominees with those nominated to represent the quantitative perspective on each topic. After lengthy discussion and negotiation the Committee produced a list of invitees; on the qualitative side, some of these turned out to be card-carrying anthropologists, and some did not. While I am personally quite satisfied with the final list, I should add that in my judgment there is an *impressive* number of other qualitative researchers "out there"—some senior and some not so senior—who would also have turned in fine performances. Participation in the selection process dramatically reinforced my earlier conviction that ethnography-applied-to-education as a professional subfield has truly come of age.

The Monterey Workshop itself was well organized, thanks to the interpersonal and organizational skills of William and Betty, and Marion Lentz. In terms of sheer *quantity* of words, we communicated energetically. In terms of the *quality* of communication, however, we sometimes fell short. We tended, predictably enough, to fall into two moeities, the metricians and the ethnographers, and into a "we-they" psychological set. I felt more than a few twitches of anthropological guilt when I sometimes discovered myself trying to "convert" the "other side" in situations where I should instead have been using the ethnographer's listening and empathizing skills to discover *common ground* that could be shared by *both* moeities. Cross-cutting the inter-moeital communication were the comments of several policy-makers and practitioners, whose participation should be acknowledged with thanks. An additional word about "common ground" is advisable. To me, what is vital is not at all that we seek to evolve into a science where quantitative and qualitative data will in all cases be mutually translatable one into the other, in some total sense. To me, "common ground" means that increasingly we will develop both the skilled scientific personnel and the procedural rules whereby we can, as a matter of consensus, decide, in particular cases, whether a given bit of qualitative data and a given bit of quantitative data are *convergent*, or otherwise, in the conclusions they lead to. Having said that, I hasten to add that where "translatability" *is* methodologically possible, I believe that it is, in general, scientifically desirable. And the direction of the translatability will doubtless generally be from qualitative to quantitative—as, for example, in the use of the Likert or Guttman Scales. However, I frankly doubt that, fifty or even a hundred years hence, we shall have arrived at anything like complete translatability, and I am old fashioned enough that I almost prefer that this be the case. I find myself hoping that it will always be true that an intuitive holistically and sometimes humanistically oriented approach to educational phenomena will enjoy a respected place in our collective armamentarium of methods. But such an approach should, and will, command more respect if it is *teamed with* a quantitative approach to those variables in the situation that lend themselves to such approach. And if we can manage to train the next generation of researchers so that both skills are lodged in the same skull, so much the better.

It is natural for the anthropologists who read this work to speculate as to what the Workshop contributed or symbolized as far as the historical development of the field of Anthropology and Education is concerned. John Herzog has hazarded the prediction that the Workshop might in the future come to be seen as the third major milestone in the maturation of the field—the first being the Stanford Conference of 1954 organized by George D. Spindler, and the second being the Miami Conference of 1968 organized by Fred O. Gearing. (See: George D. Spindler, Ed., *Education and Anthropology*, Stanford University Press, 1954; and Murray L. Wax, Stanley Diamond, and Fred O. Gearing, Eds., *Anthropological Perspectives on Education*, Basic Books, 1971.) Whether the Workshop deserves such a lofty place in our field's relatively short history is a matter for each reader to judge. My own guess is that the variance across readers' judgments will be greater, simply (or possibly) because the variation of experience among contributors and readers is so much greater.

My own response to Dr. Herzog's proposition is a bit of a cop-out. I wonder whether the Monterey Workshop was sufficiently similar to the Stanford and Miami Conferences to warrant comparison. The differences are important. The first two conferences concerned *scope* and *theory* questions more than methodological ones, and were organized and led by anthropologists, with non-anthropologists distinctly in a minority. The Monterey Workshop, by contrast, was basically organized by scholars whose early intellectual roots were elsewhere than in anthropology/ethnography, but who, to their delight, had discovered for themselves that ethnography could yield rich results. In the Stanford and Miami instances, anthropologists were essentially organizing *themselves* for the educational adventure; at Monterey, educational researchers were, in a way, *inviting* anthropologists to share. I think this difference betokens some kind of sea change in the way (non-anthropological) educational researchers perceive ethnography. This perception is, understandably, not as broad or elaborated as that of anthropologists—yet it is a growingly positive one, and part of a broader and accelerating process toward rapprochement between the quantitative and qualitative traditions in U.S. social science research in general. The mere fact that the Monterey Workshop *symbolizes* the CAE's involvement in this great rapprochement should be a source of satisfaction to many of us.

In a sense, however, it is not so important that we *judge* the importance of the Workshop as that we *predict* it: how we look back upon the Workshop at some future date depends not only on the quality of the presentations as of A.D. 1976, but upon what we, as a set of related professions, will have actually *done* in the interim, to follow up on the initiatives taken in Monterey—upon what, in the interim, we will have done in proceeding to *invent* our own future.

As we proceed to invent our future, what are some of the ways in which the leads developed at the Monterey Workshop can be productively pursued so that educational research as a whole will move forward, and so that the field of Anthropology and Education will develop fruitfully? Rather than comment specifically on the many rich insights developed in various of the individual articles that follow (which space does not permit), I will pose below five broad themes that you might find useful to bear in mind as you read through this work. The themes are purely *suggestive* and are not intended to constitute a complete or logically ordered agenda.

1. **Cultural Variablility and Contextualization.** For the most fertile synthesis of quantitative and qualitative methods to occur, it is necessary, among other things, that the broadest possible range of cultural contexts be included in our thinking. One (understandable) limitation of the Monterey Workshop is that only a few distinctive cultures were seriously discussed, and they tended to be cultures found within the boundaries of the United States. The contextualization of our discussions was based on *American* educational

problems, rather than on educational phenomena in *various* cultures, which would be the usual practice in anthropology. For this reason, when conferees discussed "qualitative" analysis, they tended, in effect, to refer to *intra*-cultural qualitative distinctions, rather than *inter*-cultural distinctions or comparisons. I believe it is vital for educational research in America (and for the Anthropology and Education portion thereof) to *globalize* its research scope. There are whole types of culture (such as peasant culture) which are seriously under-represented within the borders of the U.S.

2. **Inclusion of Quantitative and Qualitative Variables in the Same Model**. Quantitative model builders have developed various ways (e.g., through the use of dummy variables) which permit qualitative variables to be included in a basically quantitative model. Full exploitation and development of such strategems will help educate those whose bias is qualitative, to various ways in which qualitative and quantitative variables can be included in the *same* model, so that the relative explanatory power of each variable can be ascertained, regardless of whether that variable is qualitative or quantitative.
3. **Question Delineation**. It is in the nature of things that quantitative researchers tend to select questions that seem to them to be evident and paramount, and then to answer those questions in ways that are precise and verifiable. Qualitative researchers, on the other hand, not only take much longer in deciding, with any finality, what questions to ask, but are also generally less skilled in insuring precision and verifiability of findings. Nonetheless, qualitative researchers do, I think, have a substantial contribution to make in the matter of question delineation. Their very slowness in deciding what are the key questions springs from a holistic understanding (one hopes) of the overall historical, cultural, and social context within which the questions are to be asked—and qualitative researchers can often give persuasive and instructive validity-relevant reasons for their hesitation. The quantitative researchers, on the other hand, have superior offerings in such matters as the elimination of redundancy in questioning—as is accomplished when one "purifies" a scale, for example. *Joint* exploration of specific *ways* in which questions are formulated (and, later, *re*-formulated in the light of preliminary data analysis) would help establish common ground.
4. **Personal Involvement in the Research Setting**. Quantitative and qualitative researchers might do well to jointly try out various modes of deep personal involvement in the research setting, in quest of common-ground conclusions as to how this approach informs the process of formulating and re-formulating questions. One form of such involvement is intensive observation, and it is notable that in the American tradition both quantitative and qualitative research on education can involve intensive observation. A deeper form of personal involvement is the characteristic ethnographic approach of *participant* observation, and this is rare in quantitative research and not even common in qualitative research on American education—not as common as it could be. The Monterey Workshop did not, I felt, deal adequately with the whole matter of participant observation, and the empathy and informed intuition that can flow therefrom. To be sure, not every researcher wishes to, or can, take the role of the first-grader or the twelfth-grader in a school under study—yet I feel that more could be done than has been done, especially since the student role is not the only one open to the participant observer. In any case, I feel that in the future it would be well to encourage more *joint* research by quantitative and qualitative researchers using various forms of active and perduring involvement by the researchers in the on-going educational scene. The questions that ultimately get asked as a result of such involvement would almost certainly be different in some respects from those that are asked in the absence of any such involvement, and an understanding of the dynamics of such joint involvement would enrich our understanding of how quantitative and qualitative researchers can collaborate—or can*not* collaborate.
5. **The Logic of Generalization**. It is one of the strengths of the quantitative approach that it stresses the precise degree to which a given set of delimited findings may be safely generalized. The qualitative approach allows one to be less sure on this score, although its emphasis upon broad patterns of data "hanging together" and "making holistic sense" does serve as some safeguard against improper generalization. A small specialist team of quantitative and qualitative methodologists could, I think, make a contribution by working up a monograph specifically on this problem.

There is much more that could be said about the rich fare we have been given, but perhaps this will suffice. Happy reading.

INTRODUCTION

William J. Tikunoff
Beatrice A. Ward
Far West Laboratory

In their efforts to come to grips with inquiring into and understanding schooling, researchers have begun to investigate methodologies outside those of traditional educational psychology. In fact, they have been admonished by such stalwart leaders in the field as Lee Cronbach and Donald Campbell to augment quantitative data with qualitative data as well if they are to advance their science.

The papers presented here represent a singular and important effort in this direction. They were produced initially for a Workshop Exploring Qualitative/Quantitative Research Methodologies in Education held in Monterey, California in July, 1976. It is in order, then, to explain the occurrence of that workshop if the reader is to understand fully the nature of the papers themselves.

At the Far West Laboratory for Educational Research and Development we have been interested in exploring research methodologies alternative to those already used in traditional educational research modes. One such effort resulted two years ago in combining successfully qualitative and quantitative methods to produce *Special Study A: An Ethnographic Study of the Forty Classrooms of the Beginning Teacher Evaluation Study Known Sample* (Tikunoff, Berliner, Rist, 1975). Thus, we were encouraged to look at other paradigms, particularly in light of the multitude of problems confronting the educational researcher.

The workshop was the result of our interest, shared by the National Institute of Education and the Council on Anthropology and Education, to bring together experts representing the worlds of qualitative and quantitative research in order to address five of these educational problems. For each problem, therefore, we invited a paper from two recognized researchers, each working primarily in either a qualitative or quantitative mode. These were critiqued and responded to by two people who represented expertise in the problem area identified, either as researcher or as practitioner. In addition, we invited Ray Rist to present the opening paper which served to set the parameters within which we were operating. It was our hope that, given such a forum, these twenty-one people could interact productively, then rewrite their papers to incorporate any new ideas which emerged. To round out the workshop, we invited a limited number of people who attended primarily as observers.

The success of the workshop rests with whatever impact this collection will have on the field in general. However, it will be of interest to the reader to know the frustration shared by each of us as we grappled with new ideas, many of which were foreign to our experience as researchers and practitioners. Thus, it is important that a word of caution be extended to those who will read this publication.

It is not our intent to draw a philosophical line between the "qualitative" and "quantitative" perspectives and thus create a breach between these two paradigms. Indeed, paradigms for research grow out of one's experiences and discipline, and rest in the specialization that we develop. While such specialization nurtures growth in recognition of our expertise within our discipline, it also mitigates against understanding and accepting alternative paradigms. The interpretation of experience is a function of fitting a particular event into the framework of similar events in one's "experience bank." We only can perceive and understand on the basis of what we already know. This is the starting point, and teachers build upon such knowledge to construct entire concept hierarchies. We know that as we learn, we use language to label the concepts and experiences we "know," and this language forms the individual lexicons that each of us possesses. It is this lexicon that represents our bank of experiences and concepts that we bring to a learning opportunity. Thus, precisely because we perceive and interpret events differently, each of us possesses a vastly different lexicon.

It is frustrating to bring together two disparate lexicons–in this instance, those of the qualitative and quantitative researcher–to address common questions. As teachers, we might recognize the need to assess each child's lexicon so as to determine what concepts need to be taught, i.e., what words need to be added to a child's lexicon, what experiences to his/her knowledge, in order to come to an event with sufficient preparation that one could predict successful learning. As students, we understand that our task is to learn, i.e., we know that we must strive to understand in order to achieve. But as adults, too frequently we are willing to work only within the framework of our individual and/or generic lexicons.

We encourage you to keep this in mind as you read this collection of papers, for as Emile Durkheim, the French sociologist writing in the late 1800s, reminds us, science has not always existed. It is a human construct, and therefore relies on human understanding and human action to be. For Durkheim, both qualitative and quantitative aspects of educational experience are important if we are to understand that experience. It is in this spirit that we present this collection of papers.

Finally, a workshop cannot happen without the creativity and energy of people. In this vein, we wish to thank those whose efforts made this adventure possible: to John Hemphill, who initially planted the seed that became this workshop; to Virginia Koehler, National Institute of Education, for her encouragement and support as well as for financing the workshop; to members of the Advisory Committee—Robert Textor, Andy Porter, Jack Schwille—who helped formulate problem statements and suggested authors; to the Council on Anthropology and Education for co-sponsoring the workshop and publishing the papers; to the authors and respondents who contributed their ideas and energies to the task at hand; and, especially, to two who have lived with the workshop papers for six months and brought this publication to fruition: Mary Dawson for her dedication, creativity, and patience in editing these manuscripts; and to Marion Lentz for her organizational skills in coordinating the workshop and for her diligence and perseverance in producing this document.

OVERVIEW

ON THE RELATIONS AMONG EDUCATIONAL RESEARCH PARADIGMS: FROM DISDAIN TO DETENTE[1]

Ray C. Rist
National Institute of Education

> To the extent, as significant as it is incomplete, that two scientific schools disagree about what is a problem and what a solution, they will inevitably talk through each other when debating the relative merits of their respective paradigms. In the partially circular arguments that regularly result, each paradigm will be shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponent.... Since no paradigm ever solves all the problems it defines and since no two paradigms leave all the same problems unsolved, paradigm debates always involve the question: Which problems is it more significant to have solved?
>
> –Thomas S. Kuhn, The Structure of Scientific Revolutions

"Hard vs. soft." "Quantifiers vs. describers." "Scientists vs. critics." "Rigor vs. intuition." It is merely restating the obvious to suggest that the dichotomies represented by such trite cliches have too long dominated comparative discussions of varying research strategies in education. The complexities and nuances of research approaches are reduced to simple and rigid polarities. Thus the emergence of methodological provincialism reflected in the reification of the terms "qualitative methodology" and "quantitative methodology." The dialectic and interaction among all efforts to "know" or to "understand" are obscured. Further, we only hinder and cripple ourselves by a continued fixation upon what is "good" about one approach or "bad" about another. As once suggested by Homans (1949), issues of methodology are issues of strategy, not of morals.

In the quest to transform the appropriate into the orthodox, there is an inevitable distortion and skewing of the research effort. Nearly twenty years ago, C.W. Mills warned against this tendency with his castigation of those researchers who become so enamoured of one method to the exclusion of all others that they take the method as an end in itself. These researchers he terms "abstract empiricists" (Mills, 1959).

The refusal to recognize that there are different ways of "knowing" does not mean they do not exist. They do. The very fact of educational research being multiparadigmatic generates a symposium such as this. I take it to be our task here to analyze the convergent and divergent orientations inherent in our varying methodological approaches. In this way, we also may arrive at a better understanding of the possible interrelations among these differing means of approaching the social reality we all seek to comprehend.

Before moving to an analysis of these various methodologies, a short aside with regard to the title of this paper is necessary. It is my view that a situation of detente is rapidly evolving with respect to the broad categories of quantitative and qualitative research. There are at least two reasons. First, there is a general recognition among some researchers and even more practitioners that no one methodology can answer all questions and provide insights on all issues. In short, no one approach has a hegemony in educational research. Second, the internal order and logic of each approach is sufficiently articulated that it is difficult, if not impossible, to foresee the time they would merge under some broader, more eclectic research orientation.

I am not one normally to go to foreign affairs for my imagery, but I do believe that a set of accomodations is emerging whereby the various approaches, while maintaining profound tensions and different epistemological orientations, are recognizing the right of "peaceful coexistence." This coexistence both constrains and stimulates intellectual growth and development of the research efforts guided by one or another of the basic orientations. It constrains in the sense that the parameters of what is viewed as "acceptable" research are rather formal; it stimulates in that the energies of each methodology are turned inward and thus pushed towards greater refinement and sophistication (c.f. Rist, 1975).

But as with all imagery, there is some slippage between the ideal and the actual. First, there is surely the question of dominance. We are not dealing with a situation of parity among the various research methodologies. Quantitative research is *the* dominant methodology in educational research. It is more widely published, taught, accepted, and rewarded in educational research circles than any other approach. In the extreme, quantitative research is characterized as equivalent to "The Scientific Method." For example, in their widely used methodological primer, Campbell and Stanley (1963:3) term this methodological orientation "the only available route to cumulative progress." Having taken this view of quantitative research methods, it becomes understandable why those who posit an alternative set of assumptions and principles for educational research are frequently disparaged as employing an effort less than that exalted by the canons of scientific inquiry, i.e., the scientific method.

Second, there is the possibility that neither approach does, in fact, see it to be in its own best interest to pursue a policy of detente. This would be for the simple reason that neither orientation believes it particularly relevant whether any other exists or not. That is, we may have a situation in which the internal structure and principles are so self-contained and so nonreliant on external influences that the presence of other orientations is superfluous. I do not believe this to be the case, but it does remain a distinct possibility.

Research Paradigms in Education

Given that current research efforts in education are paradigmatic, it is well to spell them out in more detail prior to any comparative analysis. Building upon the work of Kuhn, Patton (1975:9) defines a paradigm in these terms:

> A paradigm is a world view, a general perspective, a way of breaking down the complexity of the real world. As such, paradigms are deeply embedded in the socialization of adherents and practitioners telling them what is important, what is legitimate, what is reasonable. Paradigms are normative; they tell the practitioner what to do without the necessity of long existential or epistemological consideration. But it is this aspect of a pardigm that constitutes both its strength *and* its weakness—its strength in that it makes action possible, its weakness in that the very reason for action is hidden in the unquestioned assumptions of the paradigm.

It is important to ferret out these "unquestioned assumptions" and subject them to examination before one attempts to assess the relative contributions of various research strategies. This is so because ultimately, the issue is not research strategies, *per se*. Rather, the adherence to one paradigm as opposed to another predisposes one to view the world and the events within it in profoundly differing ways (cf. Becker, 1967; Gouldner, 1970). The power and pull of a paradigm is more than simply a methodological orientation. It is a means by which to grasp reality and give it meaning and predictability. As Kuhn (1970:46) has suggested:

> That scientists do not usually ask or debate what makes a particular problem or solution legitimate tempts us to suppose that, at least intuitively, they know the answer. But it may only indicate that neither the question nor the answers are felt to be relevant to their research. Paradigms may be prior to, more binding, and more complete than any set of rules for research that could be unequivocally abstracted from them.

If paradigms do, in fact, constitute more than a "set of rules for research," then it is necessary to elaborate upon the ways that they do. In this way, the research orientations are themselves grounded in a perspective beyond simple questions of methodological procedure.

When we speak of "quantitative" or "qualitative" methodologies, we are, in the final analysis, speaking of an interrelated set of assumptions about the social world which are philosophical, ideological, and epistemological. They encompass more than simply data gathering techniques.

To assume otherwise about the nature of methodology is to imply that it is "atheoretical," suitable for valid scientific use by any knowledgeable user. On the contrary, the selection of a particular methodology is profoundly theoretical, regardless of its relative availability. Research methods represent different means of acting upon the environment. To choose one line of action over and against another is to have foregone others available from a different perspective and orientation. Each method reveals peculiar elements of symbolic reality. And to accentuate one aspect of that reality vs. another is to influence both observations and conclusions (Denzin, 1970:298). All knowledge is social. The methods one employs to articulate knowledge of reality necessarily flow from beliefs and values one holds about the very nature of that reality.[2] In personalistic terms, I believe this same point can be made, for example, by comparing the methods of classroom observation represented by Ned Flanders and Jules Henry, or Jane Stallings and Philip Jackson.

Recognizing full well that I may be guilty of the same reification of orientations that I criticized earlier, I would nevertheless like to pursue an assessment of the quantitative and qualitative approaches by placing them in juxtaposition. Creating this dichotomy is done with the aim of capturing the underlying and fundamental elements in each paradigm. The strategy here will be twofold: first, a very brief set of comments about the epistemological nature of each methodology and, second, a comparison of several dominant motifs and patterns that serve to clarify the alternative emphases inherent in each approach.

Quantitative Orientations

Quantitative methodologies assume the possibility, desirability, and even necessity of applying some underlying empirical standard to social phenomena. Based on these premises, there has arisen a concerted and widespread effort to formally test nomothetic propositions. Such research is assumed to contribute towards creating enduring theoretical structures. In fact, Suppes (1974) suggests that theorizing on the basis of such data collection procedures becomes the principal duty of researchers and that in due course, those who follow in the footsteps will erect "theoretical palaces" on the foundations now being laid.

Quantitative research holds to a view that the progression of knowledge moves on a continuum from observation to experimentation to theoretical development. I believe it is safe to say that the emphasis has been the latter linkage, between experimentation and theoretical development, as opposed to the former, between observation and experimentation.[3] This may be the result, at least in part, of the fact that for the quantitative researcher, working at the level of inductive statistics is intrinsically more interesting than working with descriptive statistics (cf. Blalock, 1960:4). From this orientation, it is less challenging and less creative to describe than to infer and induce properties of a population on the basis of known sample results. As Blalock notes (1960:5):

> Statistical inference, as the process is called, involves much more complex reasoning than does descriptive statistics, but when properly used and understood becomes a very important tool in the development of a scientific discipline. Inductive statistics is based directly on probability theory, a branch of mathematics.

But aside from whether one statistical approach is more challenging and creative than another, there remains for the quantitative researcher the belief that knowledge is cumulative and that the verification of what is known through experimentation is central to the scientific endeavor. As Campbell and Stanley (1963:2) have suggested regarding experimentation, it is "the only means for settling disputes regarding educational practice, the only way of verifying educational improvements, and the only way of establishing a cumulative tradition in which improvements can be introduced without the danger of a faddish discard of old wisdom in favor of inferior novelties."

Stated in this way, the paradigm governing quantitative methodologies is one derived from the natural sciences. Human events are assumed to be lawful; man and his creations are part of the natural world. The development, elaboration, and verification of generalizations about that natural world become the first task of the researcher. From that one aspires to amass empirical generalizations; then to refine and restructure them into more general laws; and finally to weave these scattered and disparate laws into coherent nomothetic theory. In short, efforts are predicated upon a belief in the correctness of the scientific method as it is practiced in the natural sciences.[4, 5]

Qualitative Orientations

The epistemological questions raised by qualitative methodology challenge the presuppositions of the natural science approach to scientific investigation. Whereas the latter may assume that the study of observable deeds and expressed words is adequate to produce knowlege about man and his natural world, qualitative methodologies assume there is value to an analysis of both the inner and outer perspective of human behavior. In the German, the term is *verstehen*. This inner perspective or "understanding" assumes that a complete and ultimately truthful analysis can only be achieved by actively participating in the life of the observed and gaining insights by means of introspection.

Emphasis is placed upon the ability of the researcher to "take the role of the other," to grasp the basic underlying assumptions of behavior through understanding the "definition of the situation" from the view of the participants, and upon the need to understand the perceptions and values given to symbols as they are manipulated by man. Qualitative research is predicated upon the assumption that this method of "inner understanding" enables a comprehension of human behavior in greater depth than is possible from the study of surface behavior, the focus of quantitative methodologies. As Filstead (1970:6) has noted:

> Qualitative methodology refers to those research strategies, such as participant observation, in-depth interviewing, total participation in the activity being investigated, field work, etc., which allow the researcher to obtain first-hand knowledge about the empirical social world in question. Qualitative methodology allows the researcher to "get close to the data," thereby developing the analytical, conceptual, and categorical components of explanation from the data itself.

This view of the means by which knowledge and understanding are developed is essentially one of inductive analysis. Theory begins with an extrapolation from "grounded events." One begins not with models, hypotheses, or theorems, but rather with the understandings of frequently minute episodes or interactions that are examined for broader patterns and processes (cf. Glaser and Strauss, 1967). It is from an interpretation of the world through the perspective of the subjects that reality, meaning, and behavior are analyzed. The canons and precepts of the scientific method are seen to be insufficient; what are needed are intersubjective understandings.[6]

Having sketched in broad strokes what Gouldner (1970) would term the "domain assumptions" behind these two methodological orientations, what follows is an effort to examine several issues in more detail. Specifically, qualitative and quantitative methodologies will be assessed in terms of the polarities of reliability vs. validity, objectivity vs. subjectivity, and holistic vs. component analysis. While any number of such diads could be constructed, these three should provide a sufficient map upon which to chart the convergences and divergences of the two research paradigms in question.

Reliability vs. Validity

Implicit in much that has been said thus far is that paradigms provide the framework or boundaries within which researchers structure their inquiry. They suggest what is appropriate to study, what questions to ask, what aspects of the phenomenon to emphasize, what standards for analysis, and what forms of interpretation to apply. Thus in any comparison of qualitative and quantitative research paradigms, there is the immediate question of emphasis (cf. Myrdal, 1972:161). Succinctly, it is my view that the emphasis within quantitative methodologies on an emulation of the scientific method has led it to emphasize reliability while qualitative methodologies have emphasized validity.[7]

The very nature of quantitative research in accentuating the cumulative properties of hypothesis testing and theory building necessitates a high degree of consensus among scientists (cf. Merton, 1957:448). Or, in the terms of Thomas Kuhn, quantification is at the very heart of the paradigm of "normal science." Such "science" is not possible if there is not a high degree of replicability and consistency among findings.

But all is not harmonious or parsimonious among the quantitatively oriented researchers. An emphasis upon reliability has its limits. As Cronbach has noted in this regard (1975:124):

> The time has come to exorcise the null hypotheses. We cannot afford to pour costly data down the drain whenever effects present in the sample "fail to reach significance." ...Let the author file descriptive information, at least in an archive, instead of reporting only those selected differences and correlations that are nominally "greater than chance." Descriptions encourage us to think constructively about results from quasi-replications, whereas the dichotomy significant/non-significant implies only a hopeless inconsistency. The canon of parsimony, misinterpreted, has led us into the habit of accepting Type II errors at every turn, for the sake of holding Type I errors in check. There are more things in heaven and earth than are dreamt of in our hypotheses, and our observations should be open to them.

Or consider this quote from Deutscher (2970:33):

> We have been absorbed in measuring the amount of error which results from inconsistency among interviewers or inconsistency among items on our instruments. We concentrate on consistency without much concern with what it is we are being consistent about or whether we are consistently right or wrong. As a consequence, we may have been learning a great deal about how to pursue an incorrect course with a maximum of precision... Certainly zero reliability must result in zero validity. But the relation is not linear, since infinite perfection of reliability (zero error) may also be associated with zero validity.

When one turns to qualitative methodologies, the emphasis is quite different. Here the concern with validity is central. The researcher is encouraged to get close to the data, to develop an empathetic understanding of the observed, to be able to interpret and describe the constructions of reality as seen by the subjects, and to be able to articulate an inter-subjectivity with regard to the phenomenon being studied. As Patton (1975:19) has noted: "The overriding issue in the *verstehen* approach to science is the meaning of the scientist's observations and data, particularly its meaning for participants themselves. The constant focus is on a valid representation of what is happening...."

Ideally, both paradigms would want high reliability and high validity. But the reality of the different emphases suggests that along this continuum, one orientation is the mirror opposite of the other. And this should immediately make apparent how, in the debates over the relative merits of the two paradigms, each finds fault in the other for the absence of its own strength. Quantitative researchers castigate qualitative researchers on their lack of reliability and their lack of work towards a cumulative body of "scientific knowledge." In an effort to meet this criticism, qualitative researchers at times make an almost pathetic attempt to argue for the "inter-rater reliability" among their field observers—a defensiveness suggesting that the manner in which quantitative researchers have defined "the scientific method" does hold a powerful appeal.

Alternatively, qualitative methodologies fault the quantitative researchers for not understanding the "meanings" behind their statistical formulations. Thus the dictum, "Statistical realities do not necessarily coincide with cultural realities." A correlation on paper may, in reality, be no correlation at all. This I take to be the caution voiced by Deutscher whom I just quoted. Parenthetically, I do not find much sense of alarm or concern among quantitative researchers about this question of validity. It may well be that the pursuit of the natural science model of research is so well established and so ingrained that questions of validity take an obvious backseat to issues of reliability.

Subjectivity vs. Objectivity

In the debate among those of the two paradigmatic persuasions, perhaps nowhere are nerves rubbed more raw than in the assessment of subjectivity vs. objectivity. While objectivity is considered the *sine qua non* of quantitative methodologies, qualitative approaches emphasize the need for *verstehen* or a subjective interpretation of the social phenomena in question. Having stated the dichotomy in this manner, it is necessary immediately to say that the meanings attached to these terms have been constantly confused, and the perspective that extols the one is used to condemn the other.

But following the lead of Scriven (1972:94-95), I agree that quantitative methods are no more synonymous with what we assume when we use the term "objectivity" than are qualitative methods synonymous with what we assume coincides with the term "subjectivity." As Scriven suggests: "Errors like this are too simple to be explicit. They are inferred confusions in the ideological foundations of research, its interpretations, its applications."

Attempting to ferret out the confusions in understanding, Scriven (p. 95) provides the following definitions:

> The terms "objective" and "subjective" are always held to be contrasting, but they are widely used to refer to two

> quite different contrasts, which I shall refer to as the *quantitative* and *qualitative* senses. In the first of these contrasts, "subjective" refers to what concerns or occurs to the individual subject and his experiences, qualities, and disposition, while "objective" refers to what a number of subjects or judges experience—in short, to phenomena in the public domain. The difference is simply the *number* of people to whom reference is made, hence the term "quantitative." In the second of the two uses, there is a reference to the *quality* of the testimony or to the report or the (putative) evidence, and so I call this the "qualitative" sense. Here "subjective" means unreliable, biased, or probably biased, a matter of opinion, and "objective" means reliable, factual, confirmable, or confirmed, and so forth.

It is in the second sense, in the "quality" of the report, that the tension between qualitative and quantitative methodologies becomes heated. It is precisely to avoid the fate of unreliable, biased, or opinionated data that reliability is stressed in quantitative approaches. But for the same goal, qualitative researchers will seek validity through personalized, intimate understandings of the social phenomena, stressing "close in" observations to achieve "factual, reliable, and confirmable" data. Having said this, we come full circle to the first part of Scriven's set of definitions. For at this point, the quantitative methodologist would pursue confirmation through the use of a number of subjects, while the qualitative methodologist might undertake an intensive case study of a small group or even some particular individuals. We are back to a reconfirmation of the view that the very basis by which to confirm or dispute, to accept or reject, to "know," are paradigm dependent.

Scriven's 1972 article is entitled "Objectivity and Subjectivity in Educational Research." I find it an important contribution to the effort to detach the traditional connotation of "subjectivity" from qualitative research and "objectivity" from quantitative research. Scriven has argued that instead there are two basic components to any scientific endeavor—prediction and understanding. Prediction, of course, has long been accepted as a goal of the scientific effort, though in its reified form, it has been reduced to simply an assessment of reliability. When he turns to the role of understanding in science, Scriven notes (1972:127):

> ...Understanding, properly conceived, is in fact an "objective" state of the mind or brain and can be tested quite objectively; and it is a functional and crucial state of the mind, betokening the presence of skills and states that are necessary for survival in the sea of information. There is nothing wrong with saying, in this case, that we have simply developed an enlightened form of inter-subjectivism. But one might also equally well say that we have developed an enlightened form of subjectivism—put flesh on the bones of empathy.

I agree here with Patton (1975:22) that the strength of Scriven's analysis lies in his suggesting that the notion of dual perspectives goes to the very heart of the tension between the quantitative and qualitative paradigms. For in the final analysis, such a perspective suggests that two researchers, working from different theoretical assumptions and different methodological orientations, may literally not see the same phenomenon, though involved in simultaneous observation. Or as Kuhn has suggested (1970:113), "something like a paradigm is necessary to perception itself." In only a slightly different context, the same issue is spoken to by Smith and Geoffrey (1968:255) in their comments on what they termed the "two realities problem."

It is one thing to recognize these differences in the basis of analysis and interpretation; it is another to set them in concrete and declare a cold war. The continued disdain implied by the selective and pejorative use of the terms "objective" and "subjective" when speaking of alternative methodological approaches does damage far beyond any reasonable intellectual clarity they might provide. And the rubble generated by such acrimony only gets in the way of our work on the question posed by Kuhn at the beginning of this paper, "Which problem is it more significant to have solved?"

Component vs. Holistic Analysis

Understandings of causality are at the heart of the scientific endeavor. Whether this pursuit of knowledge is for its own sake or to establish a basis from which to intervene to modify current conditions, the articulation of cause and effect relations is of the utmost priority. And once again, in a comparison of quantitative and qualitative methodologies, there are basic differences in how the analysis of causality is undertaken. The manner in which the topic of investigation is defined, the modes of data collection, the means of analysis, and the presentation of findings all diverge between these two paradigmatic approaches for the study of causal relations (cf. Rist, 1977: forthcoming). Neither, of course, represents an omnibus strategy for all assessments of causality, but it is apparent that within each framework rather elaborate strategies do exist.

Within the quantitative orientation, the emphasis upon the ability to manipulate variables is critical for the reason that such manipulation is central to experimentation. And as noted earlier in the quote from Campbell and Stanley, experimentation is the final arbiter of educational practice, educational improvements, and the cumulation of educational knowledge. Thus the rationale for the large number of experimental studies with a defined set of variables, one of which is the treatment variable, and the effort to separate out cause and effect. In fact, the very names of the statistical methodologies used in the assessment of these cause-effect relations gives evidence of the emphasis upon component analysis—multiple regression analysis, partial correlation analysis, linear regression analysis, nonlinear regressional analysis, correlation matrix analysis, etc.

Patton (1975:29) has nicely commented upon this relation of experimentation and educational research:

> Treatments in educational research are usually some type of new hardware, a specific curriculum innovation, variations in class size, or some specific type of teaching style. One of the major problems in experimental educational research is clear specification of what the treatment actually is, which infers controlling all other possible causal variables and the corresponding problem of multiple treatment interference and interaction effects. It is the constraints posed by controlling the specific treatment under study that necessitates simplifying and breaking down the totality of reality into small component parts. A great deal of the scientific enterprise revolves around this process of simplifying the complexity of reality.

The rationale used by quantitative methodologists for employing component analysis is stood on its head, so far as qualitative methodologist are concerned. From their perspective, it is precisely because reality cannot be broken down into component parts without the severe risk of distortion that a holistic analysis is necessary. Focusing on a narrow set of variables necessarily sets up a filtering screen between the researcher and the phenomena he is attempting to comprehend. Such barriers, from the vantage point of those employing a holistic analysis, inhibit and thwart the observer from a necessary closeness to the data, from an understanding of what is unique as well as what is generalizable from the data, and from perceiving the processes involved in contrast to simply the outcomes.

The reactions among some qualitative researchers to the extreme emphasis upon component analysis to the virtual exclusion of holistic analysis in our studies of American education have been strident. Consider this comment by Deutscher (1970:33) on the use of component analysis in the evaluation of educational programs:

> We knew that human behavior was rarely if ever directed, influenced or explained by an isolated variable; we knew that it was impossible to assume that any set of such variables was addictive (with or without weighing) we knew that the complex mathematics of the interaction among any set of variables, much less their interaction with internal variables, was incomprehensible to us. In effect, though we knew they did not exist, we defined them into being.

To reiterate, there is no omnibus strategy for our study of causality. Rather, what appears more realistic is to assume that different methodological approaches are appropriate for different levels of analysis and for different levels of abstraction. The methodology should follow the answering of the questions of for whom and for what ends the analysis is being undertaken (cf. Broadhead and Rist, 1976). Regardless of the methods employed, the assessment of any causal relation should be for reasons of its being important, not simply because it can be done. The very parsimony of saying that the method should match the problem, however, may hide as much as it elucidates. For if the analysis of this paper is correct, then stating the problem, giving it definition and form, as well as selecting the appropriate methodological techniques for its analysis are all the result of the paradigmatic spectacles one sees fit to wear.

I do not want to carry this imagery much further, but if we are serious about our quest for an understanding of the social reality about us and the causal relations within it, then what may be most needed are researchers capable of wearing bi-focal or even tri-focal lenses. In this regard, I am particularly impressed with the sensitivity demonstrated by Shapiro in her evaluations of innovative Follow Through classrooms. She seems well to have sensed the nuances of classroom life that necessitated a combination of qualitative and quantitative methodologies to achieve an accurate portrayal of the impact of Follow Through. Consider, for example, these comments (Shapiro, 1973:541):

> The relevance and appropriateness of the classroom and the test situation as locations for studying the impact of schooling on children requires reevaluation. Each can supply useful information, but in both situations the evidence is situation-bound. Neither yields pure measures, and it is necessary to consider the type of school situation the children are in and the developmental status, as well as the social and sociological factors that determine or have determined the children's expectations, perceptions, and styles of thinking and communication with other children and adults. What may be an appropriate situation for assessing some groups may lead to misevaluation of others. ...It is an old chestnut that psychological dimensions cannot be defined in terms of their physical equivalence: psychologists who are trying to study the impact of different kinds of experience on different kinds of children must be able to shift their expectations and tools depending upon the contexts in which they are working.

Conclusions

There are several rather straightforward conclusions to be drawn from the preceding analysis.

First, if in fact we do find ourselves in a situation of multiple paradigmatic perspectives on educational research, then it is not appropriate to think in the near future of there being a "grand synthesis" of quantitative and qualitative methodologies. If the two major paradigms do exist as outlined here, each with its own internal order and logic, and neither finds its present framework for analysis unsuitable, they will continue to prosper. It would only be when one or another of the approaches no longer believed in the utility and appropriateness of its paradigm that new syntheses might become possible. This may already by happening on the fringes of each paradigm, but surely not at the center. In this light, the spirit of detente may be the most we should anticipate.

Second, the fact that these two paradigms are in tension over the very most basic assumptions upon which they base their research efforts opens up the potential for a dialectic where the resolution is not an "either/or" but each answering a part of the question at hand. If each approach does provide a perspective which tends to be the mirror opposite of the other, the creative effort becomes one of finding ways to take these partial images of reality and piece them into a new orientation or perspective.[5] It may well be that some of the most intellectually stimulating and exiting developments in educational research over the next decade will be in working out the implications of the dialectic. If breakthroughs are to come, they will happen, as Kuhn (1970:110) suggests, when "scientists see new and different things when looking with familiar instruments in places they have looked before." It may well be that when the "familiar instruments" of quantitative and qualitative methodologies are juxtaposed, we will "see new and different things."

Third, and a paradox in light of the second point, is that with these two paradigms moving in their own spheres and with their own rules of evidence and acceptability in their respective communities, we confront one more example of the phenomenon of contemporary research leading to divergences rather than convergences. As each methodology is now more sophisticated than ever, as basic concepts are overhauled and refined, as new distinctions formulated, and as the sheer amount of research evidence continues to grow, we find new arguments and complications rather than new answers and resolutions. Speaking on this issue as it relates specifically to social policy research, Cohen and Weiss (1976) have noted:

> The improvement of research on social policy does not lead to greater clarity about what to think or what to do. Instead, it usually tends to produce a greater sense of complexity. This result is endemic to the research process. For what researchers understand by improvement in their craft leads not to greater consensus about research problems, methods and interpretations of results, but to more variety in the ways problems are seen, more divergence in the ways studies are carried out, and more controversy in the ways results are interpreted. It leads also to a more complicated view of problems and solutions, for the progress of research tends to reveal the inadequacy of accepted ideas about solving problems. The ensuing complexity and confusion are naturally a terrific frustration both to researchers who think they should matter and to officials who think they need help.

If Cohen and Weiss are accurate in their assessment, their comments suggest that a situation of multiple visions and understandings of reality is unescapable. And the task still remains of how then to piece our collage of realities together. Which leads to my fourth and final point.

We suffer for the lack of appropriate language and conceptual frameworks for locating both paradigms in a relation to one another. I am not sure we would recognize the collage even if we saw it. And one consequence among many of this lack of coherent organizing principles is that we probably will have to reconcile ourselves to a number of ultimately fruitless endeavors and wasted deadends. As we set out to explore these tangled and complex multiple realities with tangled and complex methodologies, the odds appear stacked against us.

But I suspect for many of us there remains the vision of developing a means to comprehend the diversities and nuances of the educational experience. And if we can come to comprehend it, then perhaps we will find the will to transform it. To learn of the ways in which to make learning and schooling both stimulating and exciting experiences for children would be no mean feat. And there are few other tasks more worthy of our efforts.

Notes

1. I wish to acknowledge the fruitful comments from Harold L. Hodgkinson on this topic. Our discussion sharpened for me several key issues raised in this paper.
2. There is yet a further philosophical issue here as well. Not only does the use of one methodological approach as opposed to another change the means by which one perceives the reality under study, but also the very reality to which a researcher has applied a method is itself continually in a state of change. As all knowledge is social, so also all reality is social. To wait for absolutes is to wait for Godot. Social systems are ongoing, regardless of how stable they may appear. Put differently, no methodology allows us to step twice in the same stream in the same place.
3. Cf. this quote from Cronbach (1975:124): "Originally, the psychologist saw his role as the scientific observation of human behavior. When hypothesis testing became paramount, observation was neglected, and even actively discouraged by editorial policies of journals. Some authors now report nothing save F ratios."
4. I raise this only as an aside, but I find it of interest that, to my knowledge, there has been no systematic tracing out of the manner in which natural science methods have been brought over into the social and behavioral sciences. Are there adaptations and mutations in the transfer process? What aspects of natural science methodology are relevant? Which are not? Are those branches of the natural sciences which are not experimental in nature (astronomy and geology, for example) able to contribute to our methodological sophistication? The analysis necessary for the answers is in the domain of the sociology of knowledge. And in the absence of such answers, I wonder if we are not at times a bit hasty to accept the "natural science" model as, in fact, the one from which current quantitative approaches have come.
5. For a more elaborate and more complete analysis of the

epistemological underpinnings between the scientific method and the natural sciences, I suggest the following sources which I found extremely beneficial: Thomas Kuhn, *The Structure of Scientific Revolutions;* Ernest Nagel, *The Structure of Science;* and Abraham Kaplan, *The Conduct of Inquiry.*

6. To suggest several citations which provide the epistemological underpinnings for the use of qualitative methodologies, I would offer the following: Alfred Schutz, *The Phenomenology of the Social World;* Herbert Blumer, *Symbolic Interactionism;* and George H. Mead, *Mind, Self, and Society.*
7. I find Patton's summary (1975) of these two concepts quite sufficient. "Reliability concerns the replicability and consistency of scientific findings." One is particularly concerned here with inter-rate, inter-item, interviewer, observer, and instrument reliability. Validity, on the other hand, concerns the meaning and meaningfulness of the data collected and instrumentation employed. Does the instrument measure what it purports to measure? Does the data mean what we think it means?
8. We may have one promising example at hand of the potential for a creative breakthrough once two paradigms are placed in a dialectic with one another. I am referring to the strides we have made in the heredity-environment debate over individual intelligence. So long as each existed without having to account for the other, little progress was made. But after a period of attempting to grasp the contributions of each in relation to its alternative, new insights are flourishing and promising research avenues opening up.

References

Becker, H.S. "Whose Side Are We On?" *Social Problems* 14:239-47, 1967.

Blalock, H.M. *Social Statistics.* New York: McGraw Hill, 1960.

Broadhead, R.A., and R.C. Rist. "Gatekeepers and the Social Control of Social Research." *Social Problems* 23:325-26, 1976.

Campbell, D.T., and J.C. Stanley. *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally, 1963.

Cohen, D., and J. Weiss. "Social Science and Social Policy: Schools and Race." Paper presented to the International Symposium on "Education, Social Science and the Judicial Process," National Institute of Education, Washington, D.C., February, 1976.

Cronbach, L.J. "Beyond the Two Disciplines of Scientific Psychology." *American Psychologist* 30:116-127, 1975.

Denzin, N.K. *The Research Act: Introduction to Sociological Methods.* Chicago: Aldine, 1970.

Deutscher, I. "Words and Deeds: Social Science and Social Policy." In W.J. Filstead (ed.). *Qualitative Methodology.* Chicago: Markham, 1970.

Filstead, W.J. *Qualitative Methodology.* Chicago: Markham, 1970.

Glaser, B., and A. Strauss. *The Discovery of Grounded Theory.* Chicago: Aldine, 1967.

Gouldner, A. *The Coming Crisis in Western Sociology.* New York: Basic Books, 1970.

Homans, G. "The Strategy of Industrial Sociology." *American Journal of Sociology* 54:330-337, 1949.

Kuhn, T.S. *The Origins of Scientific Revolutions.* Chicago: University of Chicago Press, 1970.

Merton, R.S. *Social Theory and Social Structure.* Glencoe (Ill.): Free Press, 1957.

Mills, C.W. *The Sociological Imagination.* New York: Oxford, 1959.

Myrdal, G. "How Scientific Are the Social Sciences?" *Journal of Social Issues* 28:151-70, 1972.

Patton, M.Q. *Alternative Evaluation Research Paradigm.* Grand Forks: University of North Dakota Press, 1975.

Rist, R.C. "Race, Policy, and Schooling." In I.L. Horowitz (ed.). *The Use and Abuse of Social Science.* New Brunswick (N.J.): Trans-Action Books, 1975.

__________ *The Invisible Children: A Study of School Desegregation in American Society.* Cambridge: Harvard University Press, 1977 (forthcoming).

Scriven, M. "Objectivity and Subjectivity in Educational Research." In H.B. Dunkel, et al. (eds.). *Philosophical Redirection of Educational Research.* Chicago: National Society for the Study of Education, 1972.

Shapiro, E. "Educational Evaluation: Rethinking the Criteria of Competence." *School Review* 81:523-49, 1973.

Smith, L. and W. Geoffrey. *The Complexities of an Urban Classroom.* New York: Holt, Rinehart, and Winston, 1968.

Suppes, P. "The Place of Theory in Educational Research." *Educational Researcher* 3(6):3-10, 1974.

NEXT STEPS IN QUALITATIVE DATA COLLECTION

Educational research has long attended to quantifying dependent variables as a way of describing learning outcomes. It is in the domain of the independent variable, or the interactions and characteristics of the classroom and the teaching-learning participants as a casual factor, that we have been weakest. Techniques are emerging which can help to identify independent variables. What are some of these, and how can they be applied to gathering the qualitative data important to the identifications of such variables?

THE COLLECTION AND ANALYSIS OF ETHNOGRAPHIC DATA IN EDUCATIONAL RESEARCH

Stephen E. Fienberg
School of Statistics
University of Minnesota

When I was first asked to prepare a paper for this symposium, I was uncertain what contribution I could make toward resolving the apparently "hot controversy" of qualitative versus quantitative research methodology in education. Reluctantly, I agreed to participate, and the more I read about the controversy and ethnographical educational research, the less sure I was what focus the paper should take.

My first impression of the ethnographic methodology literature was that it was antistatistical or at best astatistical. This impression was fostered by studies such as Philips' (1972) description of participant learning structures for children from the Warm Springs Indian Reservation, which contains not a single numerical summary statistic or tabular display. In a second reading, however, I began to notice the models Philips had constructed regarding communication and interaction of teachers with children. Although not mathematical or statistical models of the sort to which I am accustomed, they were models nonetheless, and I realized that the principles and methods of scientific inquiry rising from the ethnographical approach do not really differ from those used in the "psychostatistical" approach.

Thus I was led to the basic theme of this paper: that from a scientific viewpoint, there is no fundamental difference between the two sides of the qualitative/quantitative controversy (or at least should not be). It follows from this position that the process of statistical inference is basically the same for both types of research, despite comments to the contrary by such authors as Lutz and Ramsey (1974). Given these basic premises, the issues associated with the collection and analysis of ethnographic data are basically methodological and, at least from the vantage point of the statistician, not necessarily unique to ethnographic research methodology. These are positions on which I shall elaborate during this paper.

I have tried (without total success) to avoid using the qualitative/quantitative distinction posed by the symposium title. Instead, I have used the term "ethnographic" in place of "qualitative" and "psychostatistical" in place of "quantitative," primarily because the words qualitative and quantitative have technical meanings in statistics which, although related to meanings used in this symposium, are not quite the same. The term qualitative in statistics is used to describe discrete variables whose possible values are categorical in nature (e.g., the presence or absence of an attribute); the term quantitative is used to describe continuous variables that can take any value in a predefined range. It is true that much ethnographic data is qualitative in the statistical sense while much psychostatistical data is quantitative. Nevertheless, some ethnographic data involves quantitative variables and considerable psychostatistical data is categorical in nature.

Because ethnographic data are obtained by direct observation of human activity and interaction in an ongoing naturalistic manner, it is inherently multidimensional. Attempts to analyze such data that ignore this multivariate structure are likely to run into difficulties. Thus, on the surface, it would seem that the analysis of multidimensional categorical data (e.g., Bishop, Fienberg, and Holland, 1975, or Fienberg, 1977) would find many applications in studies dealing with ethnographic data. That I have been able to find no examples of such applications is less a commentary on the ethnographer's willingness to use new statistical methods than it is a reflection of the limited scope of most ethnographic investigations. Only in the context of large-scale controlled field trials are we likely to see the techniques of multivariate analysis being used for the analysis of ethnographic data.

I would like to make one additional introductory comment. Investigators in any field tend to be unaware of parallel developments in quite unrelated areas. Statistics is an exception, primarily as a result of its wide range of application to essentially all of the sciences. (Cornfield [1975] has suggested that statistics be dubbed the "bedfellow of the sciences.") Thus from my perspective as a statistician, the qualitative/quantitative controversy in education possesses many of the same features as the controversy in medicine between the use of "clinical judgment" as typified by the work of Feinstein (1967) and the use of standard statistical methods as typified by the work of Bradford Hill (1966). Although the parallel is not quite complete, I see many of the same criticisms of the use of standard statistical methodology in Feinstein's work as I do in papers advocating the use of ethnographic and related methods such as those of Lutz and Ramsey (1974) and Snow (1974). But when push comes to shove, both the clinical judgment doctors and the ethnographic researchers in education wish to make proper inferences from data. What we statisticians need to do for both the ethnographic researcher and the medical clinician is to work on the construction of suitable statistical models for the data at hand, and then develop methods for their analysis.

Scientific Inference and the Ethnographic Method

Two eminent statisticians, Karl Pearson and Harold Jeffreys, with whom many educational researchers may not be familiar, clearly state the premises on which I base my comments in this paper.

> Now this is the peculiarity of scientific method, that when once it has become a habit of mind, that mind converts all facts whatsoever into science. The field of science is unlimited; its material is endless, every group of natural phenomena, every phase of social life, every stage of past or present development is material for science. The unity of all science consists alone in its method, not in its material. The man who classifies facts of any kind whatever, who sees their mutual relation and describes their sequences, is applying the scientific method and is a man of science. The facts may belong to the past history of mankind, to the social statistics of our great cities, to the atmosphere of the most distant stars, to the digestive organs of a worm, or to the life of a scarcely visible bacillus. It is not the facts themselves which form science, but the methods by which they are dealt with (Pearson, 1892: 16 of Everyman edition [1938]).

> No matter what the subject-matter, the fundamental principles of the method must be the same. There must be a uniform standard of validity for all hypotheses, irrespective of the subject. Different laws may hold in different subjects, but they must be tested by the same criteria; otherwise we have no guarantee that our decisions will be those warranted by the data and not merely the result of inadequate analysis or of believing what we want to believe.... If the rules [of induction applied in scientific inquiry] are not general, we shall have different standards of validity in different subjects, or different standards for one's own hypotheses and somebody else's. If the rules of themselves say anything about the world, they will make empirical statements independently of observational evidence, and thereby limit the scope of what we can find out by observation. If there are such limits, they must be inferred from observation; we must not assert them in advance (Jeffreys, 1961:7).

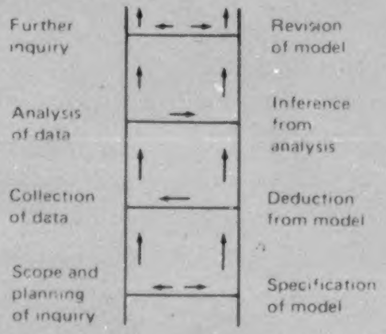
If we are to draw inferences from ethnographic educational data, we must use the same rules of inference that we use for psychostatistical educational data. The statistical models we choose may be more complicated, they may involve only qualitative variables or mixtures of qualitative and quantitative variables, and they may involve complex stochastic phenomena not representable in the form of simple systems of linear equations now commonly used in behavioral research; but the basic requirements for statistical analysis remain the same, as do the principles underlying experimentation.

The "ladder diagram" in Figure 1, adapted from Bartlett (1967), displays the sequential aspect of all scientific inquiry. Bartlett notes that this approach "permits, on the practical side, the manageable reduction of suitable data, and, on the theoretical side, the use of statistical probabilities." Statistical methods come into play at every step up the ladder, and if sensibly used, it does not really matter whether these methods be Bayesian (e.g., Novick and Jackson, 1974), classical, or some mixture of the two.

A ladder is of little use unless it is located on firm ground. Thus the scientist needs to work on a sensible problem, consider all of the relevant variables, and measure these variables in the most appropriate way. This brings us to the bottom rung. In his paper for this conference, Ray Rist quoted Filstead to the effect that "qualitative methodology allows the researcher to 'get close to the data.' " I find it astonishing that getting close to the data can be thought of as an attribute of only the ethnographic approach. Perhaps statisticians are in fact ethnographic researchers in disguise, for the good statistician working on a project as a collaborator tries to learn all about the data before designing an experiment or planning a sample survey.

FIGURE 1

Bartlett's Ladder Diagram of Scientific Inquiry



Last summer, two of my students analyzed some crime report data for the Minneapolis Police Department. The first thing I had them do, before they looked at the numbers, was to spend an evening riding in a squad car so they could observe firsthand how crime reports are really generated. This same approach applied when I worked with an ecologist analyzing data on the structural habitat of lizards, although a ten-day field trip to the Bahamas did have some nonscientific benefits! Having worked with individuals in several fields, I know that for me to design a sensible experiment involving first grade classrooms, I must observe a real classroom in action and see what happens when certain kinds of changes are instituted.

What appears to distinguish "ethnographic" from psychostatistical studies is the scope and planning of the inquiry. Rather than assess the effectiveness of teaching by the traditional techniques of test scores administered before and after some "treatment," the ethnographer chooses to investigate how events within the classroom and the interactions among teacher and students affect the learning process. This view of the basic inquiry has led ethnographers to the method of direct observation (most typically nonparticipant observation) for data collection. Through some undefined process, they analyze the resulting data and then go on to further inquiry and/or inference from the analysis.

Sometimes the ethnographer has a simple model linking the social and cultural features of the classroom to the ultimate outcomes of interest, and the data are completely consistent with the model. This is all well and good, and in such circumstances the statistician has little of value to add. More often, however, the ethnographer's conceptual model is much more complex and poorly formulated, and the data conform less clearly to the model. In these situations, methods for the analysis of data to be collected require careful consideration, and inferences from the analyses require detailed model specifications. *Now* the statistician does have something to say to the ethnographer, and without the links provided by statistical methodology between the "pole" of practice and the "pole" of theory, the ethnographer is likely to climb an endless ladder, ending up at the same place from which he started.

As I noted earlier, many ethnographical studies do involve the formulation of models. But in studies such as Philips' (1972), these models must be made explicit, perhaps even in mathematical terms, and the basic data or some summary must be presented. Only then will others be able to determine whether the inferences from the data about the model are correct. This is one of the basic lessons of scientific reporting.

The following sections deal not only with the statistical links for the ladder of the ethnographer's inquiry but also with features of the poles.

What is the Basic Analysis Unit?

Most elementary statistics textbooks state that one of the basic objectives of statistical analysis is to make inferences about a population based on information contained in an appropriately selected sample. The unit of analysis results from the type of measuring instrument used, the method of sampling, and the model of the phenomenon of interest.

Ethnographic researchers seem to have pinpointed a major flaw in much educational research: the unit of analysis need not be the same as the apparent unit of sampling. Just because tests are administered to the individual students and test scores are available at the individual level, the basic unit of analysis need not be the student. As Rist (1970) has pointed out, if we are interested in working with educational systems:

> ...there appear to be at least three levels at which analysis is warranted. The first is a macro-analysis of structural relationships where governmental regulations, federal, state, and local tax support and the presence or absence of organized political and religious pressure all affect the classroom experience.... The milieu of a particular school appears to be the second area of analysis in which one may examine facilities, pupil-teacher ratios, racial and cultural composition of the faculty and the students, ...all of which may have a direct impact in the quality as well as the quantity of education a child receives.
>
> Analysis of an individual classroom and the activities and interactions of a specific group of children with a single teacher is the third level at which there may be profitable analysis of the variations in the educational experience.

By focusing on the classroom as the unit of inquiry and analysis, the ethnographer forces us to acknowledge that even if we are interested in changes in test scores and gather information on individuals, the unit of analysis should likely be the classroom or even the school.

The same concerns have been expressed in the context of more traditional educational investigations. In comments on *Equality of Educational Opportunity* (the Coleman Report), Hanushek and Kain (1972) note: "The basic sampling units...were elementary and secondary schools attended by seven broad ethnic groups...the reader is lulled into a false sense of security by the seemingly generous sample size (569,000 students). But when it comes to school facilities, the relevant sample size is the number of schools, not the number of students." Different analyses of the Coleman Report data using different basic units lead to somewhat different results and conclusions. Those readers familiar with basic techniques in the design of experiments will not be shocked by these comments since they will recognize the nested aspects of the data in the OE Survey (i.e., students are nested within schools) and recall how certain tests in the analysis of variance of a split-plot (repeated measures) design use an error term based on main plots, while other tests are based on a sub-plot error sum of square (e.g., Snedecor and Cochran, 1967:372 or Bock, 1975: Chapter 7).

The Tyranny of Small N and the Peril of Large p

Given that the classroom is the basic unit of analysis in ethnographic educational research, what is the sample size of the typical ethnographic investigation? All too often $N = 1$, as in the well-known studies of Smith and Geoffrey (1968) and Rist (1970, 1973). Even though the information collected on a single classroom group over the period of a year or more is extremely rich, the basic fact remains that for a single classroom study, $N = 1$ no matter how many dimensions (p) the method of direct observation has allowed us to measure. The $N = 1$ case is especially troublesome because even if we measure only a single variable, we have trouble making statistical inferences from the datum. An important feature of multivariate statistical analysis is that most methods require that, at a minimum, $N \geq p$. When the variables are categorical in nature, we often require considerably larger sample sizes.

Let us briefly look at Rist's three-year study of a ghetto elementary class in St. Louis to see the problems with an $N = 1$ investigation. Rist notes that the kindergarten teacher placed children in reading groups which reflected the social class composition of the classroom as defined by such information as estimated family income (Table 1).

If we accept Rist's estimates of family income and use students as our analysis units, such a configuration would rarely occur by a random placement of students ($p < .005$ based on the usual chi-square test). Rist observes that these original seating groups persisted through the second grade, and the teachers consistently treated the Table 1 group differently from the other two, thus influencing the children's achievement.

But $N = 1$ in Rist's study. In no sense was the classroom selected at random. We are given no information about how it compares to others removed in time and place. Even if his observations for this single group of children in St. Louis are without error or bias, by what method of induction can we draw sound statistical inferences about the title of Rist's book, *The Urban School: A Factory for Failure?* A study such as Rist's may help to generate hypotheses about urban or ghetto schools; it does not allow for generalizations or broad conclusions (perhaps not even narrow ones).

TABLE 1

Distribution of Family Income by Seating Arrangement at the Three Tables in the Kindergarten Classroom

Estimated family income	Seating Arrangement		
	Table 1	Table 2	Table 3
Less than $3,000	0	4	7
Less than $12,000 but greater than $3,000	5	7	3
Greater than $12,000	4	0	0
Student totals	9	11	10

Source: Rist, 1973:88

What, then, is the ethnographer to do if he wishes to use statistical methods in making inferences from his data? There are two answers to this query, the most obvious being: increase the sample size. If N must be at least as great as p and we have a minimum number of dimensions we wish to measure, then we must increase N appropriately. This means that we must collect comparable and reliable qualitative data on each of several classrooms if we are to avoid the myriad of problems associated with missing data in multivariate analysis. The work of the Far West Laboratory addresses many of the difficulties involved in such multiple classroom ethnographic educational studies (Tikunoff, Berliner, and Rist, 1975).

Given the richness of the information collected by the ethnographer (i.e., given how large p may be), increasing the sample size may not be enough. The other direction the ethnographer must go is to build probabilistic or stochastic models for the occurrence of events and interactions over time. Such models (when valid) often lead to parsimonious descriptions of apparently complex phenomena and can have the effect of reducing the number of dimensions or parameters of interest (p) while simultaneously increasing the effective sample size, N. For example, modeling group conversations using the "who-speaks-to-whom" paradigm of F. Bales along with first or second order Markov chains allows one to go from a single classroom or $N = 1$ situation to a form of analysis where N may easily exceed 100 (see Bishop, Fienberg, and Holland, 1975: Chapters 5,8). Such stochastic modeling is especially important when dealing with longitudinal data, be it qualitative or quantitative. Beshers (1972) discusses some closely related notions of stochastic models for the educational process.

Erickson, in his paper for this symposium, describes a study involving a single kindergarten-first grade classroom where interest has been focused on participant structures that might be present in or generalizable to other contexts. The data collected would seem to be ideal for stochastic modeling, and thus he should be able to convert an $N = 1$ situation into one where the basic units of measurements are events, such as a child getting a turn to speak, of which there are large numbers.

In analyzing multidimensional data, we often make a large number of different comparisons involving individual variables and, when the data are longitudinal, involving the same variables over time. The theory of simultaneous statistical inference has shown that doing separate tests of significance or constructing separate confidence intervals for each comparison can be highly misleading (Miller, 1966). For example, the p values reported in the study by White and Watts (1973) on the development of the young child are not based on multiple comparisons for a given point in time and repeated comparisons across time; they thus suggest much sharper inferences than should be drawn from such a study.

The simultaneous inference concept goes hand in hand with observations on multiple variables. Thus the ethnographic researcher is beset on the one side by the tyranny of small N and on the other by the peril of large p. These two problems must be considered a priori, as part of the scope and planning of an inquiry.

Distinguishing Among Potential Causal Explanations

A major aim of much social science research is to provide a causal explanation of a given phenomenon, and educational research is no exception. In a randomized controlled experiment, the investigator manipulates various explanatory variables so as to assess their effects on one or more dependent or response variables. For the experiments to elucidate causal relationships, the investigator must identify in advance the important causal factors for the phenomenon in question and be capable of measuring the manifestations of the phenomenon by means of suitable response variables. Both of these aspects of a good experiment rely on substantive knowledge, yet the statistical question remains: given a large amount of observational or nonexperimental data, how can one discover possible causal factors or distinguish among potential causal explanations?

During the past fifteen years, considerable effort has gone into the development of path analysis and structural equations models (e.g., Duncan, 1975; Goldberger and Duncan, 1973), especially among economists and sociologists. But these are simply useful technical devices, and rather than expound upon them here, I would like to quote from the final chapter of Duncan's book (1975) on the topic.

> Do not undertake the study of structural equation models (or, for that matter, any other topic in sociological methods) in the hope of acquiring a technique that can be applied mechanically to a set of numerical data with the expectation that the result will automatically be "research." Over and over again, sociologists have seized upon the latest innovation in statistical method, rushed to their calculators or computers to apply it, and naively exhibited the results as if they were contributions to scientfic knowledge. The lust for "instant sociology," the superstition that it is to be achieved merely by a complication if not perfection of formal or statistical methods, and the instinct to suppose that any old set of data, tortured according to the prescribed ritual, will yield up interesting scientfic discoveries—all these pathological habits of thought are grounded (if at all) in the fallacy of induction.

To make this quote relevant in the present context one need only substitute "educational research" in place of "sociology" throughout. Duncan goes on to note that models, be they of the structural equation or some other variety, are contributions to science only if they rest on creative, substantial, and sound theory. This, of course, is totally consistent with the basic theme regarding statistical inference described earlier in this paper.

One other technique that might be useful in observational educational research for exploring possible causal relationships is the case-control study, endemic to epidemiology. In the case-control method, the investigator is usually interested in the causes of a "disease." He observes "populations" both with and without the disease and attempts to determine in what respects they differ and how these differences might be related to the disease. As with all observational studies, this approach may or may not be helpful in identifying causal relationships. The key is in the intelligent choice of "controls" with whom the individuals from the diseased population are compared. In the case-control study, the epidemiologist usually has the cases possessing the disease on hand (or at least appropriate information on them), and then selects the controls, matching on suitable socio-demographic or medical variables in order to make the controls as similar to the cases as possible, except with respect to the disease. The variables chosen for matching depend on the phenomenon of interest. Thus, in a study of cervical cancer patients one might consider age, age at first pregnancy, number of pregnancies, urban-rural status, etc., as variables for matching controls to cases.

Schneiderman and Levin (1973) point out an important shortcoming of such studies that is especially relevant to educational research: "If we match on race in a case-control study, then we are most unlikely to detect etiological factors closely associated with race. If race is highly correlated with other variables (for example, socio-economic factors) we may also lose the effects of these other variables." These cautions are similar to those found in the literature on other methods applied in nonexperimental research such as regression analysis.

Sometimes the isolation of causal factors is relatively easy, as in the following example. After an outbreak of food poisoning at a large company picnic, 304 of the 320 persons attending filled out a questionnaire about the food they had consumed. Out of all the food served, the epidemiologists ultimately focuses on potato salad and crabmeat. The resulting data are reproduced in Table 2.

How were the epidemiologists able to eliminate all the other foods from consideration? The answer is the "0" in Table 2; among those who ate neither crabmeat nor potato salad there were no cases of food poisoning. A detailed analysis of this 2 x 2 x 2 contingency table reveals that the association between illness and crabmeat is not as large as that between illness and potato salad, but the former cannot be entirely dismissed. (This analysis still does not rule out the possibility of one common causal factor common to both foods, such as mayonnaise.)

TABLE 2

Observed Three-Dimensional Data with a Random Zero

Consumer's Illness	Food Eaten			
	Crabmeat			
	Yes		No	
	Potato Salad		Potato Salad	
	Yes	No	Yes	No
Ill	120	4	22	0
Not Ill	80	31	24	23

Source: Korff, Taback, and Beard, 1952, as given in Bishop, Fienberg, and Holland, 1975:90.

Notice the importance of looking at the data at least three dimensions at a time. Had we looked only at the two-dimensional marginal tables linking each food separately to the outcome variable, we would not have had the zero to direct our inferences. While the occurrence of zeros in strategic cells in a multidimensional cross-classification may lead to the elimination of multiple causal factors, for the zero to have such force, the sample size must be sufficiently large to allow for the detection of one or more "true" causal factors.

The occurrence of strategic structural zeros is not the only reason for looking at multidimensional cross-classifications. Consider the following hypothetical example. Two mathematics teachers, Jones and Smith, have been teaching in the City school for several years, and the superintendent wants to determine who is the superior teacher. The outcome measure chosen for the evaluation is the performance of students in advanced mathematics (success or failure). The superintendent looks at the data in Table 3 and concludes that Jones is superior to Smith since 30.3% of Jones' students are "successes" whereas only 27.5% of Smith's students are.

TABLE 3

Hypothetical Data for Teacher Comparisons

Teacher	Student Performance Success	Failure	Totals
Smith	55 (27.5%)	145	200
Jones	91 (30.3%)	209	300

Before the superintendent writes his report, the school principal divides the students of Jones and Smith by race and looks at success for blacks and whites separately. The results, listed in Table 4 are a shock: for black students, Smith (5% success) is superior to Jones (1%), and for white students, Smith (50%) is again superior to Jones (45%). This is an illustration of a phenomenon known as Simpson's Paradox (Simpson, 1951) and discussed in a slightly different context by Meehl and Rosen (1955) and Lindley and Novick (1975).

TABLE 4

Further Breakdown of Hypothetical Data in Table 3

Teacher	Student Performance by Race					
	Black			White		
	Success	Failure	Totals	Success	Failure	Totals
Smith	5 (5%)	95	100	50 (50%)	50	100
Jones	1 (1%)	99	100	90 (45%)	110	200

As I have remarked elsewhere (Fienberg, 1977), Simpson's Paradox is not really a paradox at all. Rather, it is a lesson which reminds us that when we compare proportions, we must condition on all of the relevant variables. In our example, the comparison of teachers varies by the race of the student, and so it makes little sense to look at Table 3 which ignores this information. When several variables are interrelated, we must look at them all together. This, of course, is what multivariate statistical analysis is all about. However, the researcher is the one who must decide which variables to measure. If the wrong variables are measured or if important ones are omitted, the most sophisticated statistical techniques will be of little use. The message of this example is especially relevant for the design of randomized, controlled experiments. If we fail to control for variables or importance (or if we exercise faulty control), the fanciest design will produce an uninteresting result.

Next Steps: Randomized Controlled Field Trials

The major conclusion I hope you have reached from this discussion is that in addition to using multivariate methods to analyze their data, investigators need to begin thinking in terms of large-scale randomized controlled field trials (i.e., experiments). The conclusion for me is inescapable. Both the control and the randomization are necessary.

Campbell and Stanley (1963), Gilbert and Mosteller (1972), and others have noted the need for control. It plays at least two crucial roles, ensuring that the choice of "treatments" for subjects is made by the investigator, not by the subjects nor by "nature," and that at least two "treatments" are being contrasted in comparable circumstances.

But control alone is not enough. We cannot confidently compare the outcomes in two parallel situations unless we observe them both, deliberately changing the values of the causal variables from one to the other. As a result of apparently lower costs, easier execution, or other practical reasons, investigators all too often use nonrandomized and even noncontrolled trials. But even when such trials are well-executed and expensive and when time consuming evaluations are conducted, the results can be ambiguous and the interpretations conflicting. As Gilbert, Mosteller, and Light (1975) note: "Frequently the question is 'were the differences found the result of how the samples were chosen or were they due to program effects?' In several large sets of parallel studies, the results of non- randomized and randomized evaluations of the same programs conflict." The only sure way to resolve such issues is through the use of randomization, even if one's statistical philosophy is Bayesian (e.g., Rubin, 1975).

By using the term "field trial" in place of the more traditional term "experiment" I mean to convey that the study adapts to the form of the phenomenon and thus is carried out in classrooms and schools, not in the laboratory. Most of Snow's (1974) comments on representative designs for educational research are compatible with this approach, except for his willingness to accept nonrandom assignment of treatments.

Why have investigators resisted the notion of randomized controlled experiments in educational research? Gilbert and Mosteller (1972) point out that they haven't. There have been many attempts at such experiments, but few have been successful. Part of the difficulty is that many of these experiments have been small scale, modeled after psychological experiments in the laboratory (c.f. Snow, 1974). It is also the case that educational innovations often are ineffective. Of course, one of the best ways to discover such ineffectiveness is through a randomized controlled study.

Is there any example of a successful large-scale randomized controlled study in education which has found a positive effect? Gilbert, Mosteller, and Light (1975) discuss one such study in detail. A randomized controlled field study was carried out during 1971-72 under the federally funded Emergency School Assistance Program (ESAP) which sought to improve the quality of education in desegregating schools. All schools receiving funds used them for counseling, remedial programs, etc., and in addition high schools used some funds on programs related to handling problems in race relations. Because there were not enough funds for all schools, it was easy to justify the use of randomization for the allocation of funds and to imbed the study in the main ESAP program. In the South, 50 pairs of high schools and 100 pairs of elementary schools applied for ESAP grants; in each pair, one was randomly chosen to receive funds and the other not. Gilbert, Mosteller, and Light (1975) report:

> ...The major positive finding was that black males in funded high schools improved by half a grade level compared to those in high schools without the funds. Other groups were not detected as improving. The researchers suppose, but not with as strong convictions as they have for the existence of the improvement itself, that the race relations program may have influenced attitudes for the male blacks, leading to improved school performance.
>
> Because such positive findings about school performance of black males came from a randomized study, there should be relatively few disagreements about the results themselves. Thus the value of randomization in this ESAP study is great for it gave us firm inferences about a program that was adopted on a wide scale and worked.

Successful randomized controlled trials must come to grips with political and social realities. The idea that standard types of design developed for laboratory experiments can be used without change to evaluate innovations in education is clearly unrealistic. Yet examples of well-designed field trials are available. Gilbert, Light, and Mosteller point out that evaluation of medical programs suffers from many of the same problems as does evaluation of social programs, but because of doctors' diligence in evaluating the effects of their therapies, we now have several models of how to go about planning a sensible social experiment.

Fienberg, Larntz, and Reiss (1976) discuss aspects of the design of police patrol experiments that also are relevant. In fact, the parallels between the preventive patrol experiments they discuss and possible experiments involving classroom innovations are remarkable. For example, they deal at length with issues of treatment implementation (how do you get patrolmen to do what you want?), the difficulties imposed by the social system (differences among neighborhoods and the current practices of dispatching officers in response to calls for service), and the idea of using each experimental unit as its own control (since levels of crime vary greatly from beat to beat). Each of these problems has its counterparts in a typical educational research field trial involving classroom teaching. This is not to say that the design for the preventive patrol experiment could be used for a classroom study, but rather that it is possible to face up to many of the political and social obstacles to randomized controlled trials.

One of the challenges for the ethnographic educational researcher is to demostrate the superiority of the anthropological field method over the more traditional test score approach in discovering positive effects of educational innovations. There is no better way to do this than in the context of large-scale randomized controlled experiments.

References

Bartlett, M.S. "Inference and Stochastic Processes." (Presidential address.) *Journal of the Royal Statistical Society*, Series A 130:457-77, 1967.

Beshers, J.M. "Models of the Educational Process: A Sociologist's Perspective." In F. Mosteller and D.P. Moynihan (eds.). *On Equality of Educational Opportunity*. New York: Random House, 1972.

Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: M.I.T. Press, 1975.

Bock, R.D. *Multivariate Statistical Methods in Behavioral Research*. New York: McGraw-Hill, 1975.

Bradford Hill, A. *Principles of Medical Statistics* (8th ed.). New York: Oxford University Press, 1966.

Campbell, D.T., and J.C. Stanley. *Experimental and Quasi-experimental Designs for Research*. Chicago: McNally, 1963.

Cornfield, J. "A Statistician's Apology." (Presidential Address) *Journal of the American Statistical Association* 70:7-14.

Duncan, O.D. *Structural Equation Models*. New York: Academic Press, 1975.

Feinstein, A. *Clinical Judgement*. Baltimore: Williams and Wilkins, 1967.

Fienberg, S.E. *The Analysis of Cross-Classified Categorical Data*. Cambridge: M.I.T. Press, 1977.

Fienberg, S.E., K. Larntz, and A.J. Reiss, Jr. "Redesigning the Kansas City Preventive Patrol Experiment." *Evaluation Magazine* 3:124-31, 1976.

Gilbert, J.P., and F. Mosteller, F. "The Urgent Need for Experimentation." In F. Mosteller and D.P. Moynihan (eds.). *On Equality of Educational Opportunity*. New York: Random House, 1972.

Gilbert, J.P., R.J. Light, and F. Mosteller. "Assessing Social Innovations: An Empirical Base for Policy." In A.R. Lunsdaine and C.A. Bennett (eds.). *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*. New York: Academic Press, 1975.

Goldberger, A.S., and O.D. Duncan (eds.). *Structural Equation Models in the Social Sciences*. New York: Seminar Press, 1973.

Hanushek, E.A., and J.F. Kain. "On the Value of *Equality of Educational Opportunity* as a Guide to Public Policy." In F. Mosteller and D.P. Moynihan (eds.). *On Equality of Educational Opportunity*. New York: Random House, 1972.

Jeffreys, H. *Theory of Probability* (3rd ed.). London: Oxford University Press, 1961.

Korff, F.A., M.A.M. Taback, and J.H. Beard. "A Coordinated Investigation of a Food Poisoning Outbreak." *Public Health Reports* 67:909-13, 1952.

Lindley, D.V., and M.R. Novick. "Inferences from Large Data Sets to Individual Cases: Use of Appropriate Conditional Probabilities in Exchangeable Subpopulations." (Unpublished manuscript.) 1975.

Lutz, F.W., and M.A. Ramsey. "The Use of Anthropological Field Methods in Education." *Educational Researcher* 11:5-9, 1974.

Meehl, P.E., and A. Rosen. "Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores." *Psychological Bulletin* 52:194-216, 1955.

Miller, R.G. Jr. *Simultaneous Statistical Inference*. New York: McGraw-Hill, 1966.

Novick, M.R., and P.H. Jackson. *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill, 1974.

Pearson, K. *The Grammar of Science*. London: A. & C. Black, 1892.

Philips, S.U. "Participant Structures and Communicative Competence: Warm Springs Children in Community and Classroom." In C. Cazden, D. Hymes, and V. John (eds.). *Functions of Language in the Classroom*. New York: Teachers College Press, 1972.

Rist, R.C. "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education." *Harvard Educational Review* 40:411-51, 1970.

____________ *The Urban School: A Factory for Failure*. Cambridge: M.I.T. Press, 1973.

Rubin, D. "Bayesian Inference for Causality: The Importance of Randomization." Proceedings of the Social Statistics Section, American Statistical Association, Washington, D.C., 1975.

Schneiderman, M.A., and D.L. Levin. "Parallels, Convergences, and Departures in Case-Control Studies and Clinical Trials." *Cancer Research* 33:1498-1504, 1973.

Simpson, E.H. "The Interpretation of Interaction in Contingency Tables." *Journal of the Royal Statistical Society*, Series B. 13:238-41, 1951.

Smith, L.M., and W. Geoffrey. *The Complexities of an Urban Classroom*. New York: New York: Holt, Rinehart and Winston, 1968.

Snedecor, G.W., and W.G. Cochran. *Statistical Methods* (6th ed.). Ames: Iowa State University Press, 1967.

Snow, R.E. "Representative and Quasi-Representative Designs for Research on Teaching." *Review of Educational Research* 44:256-91, 1974.

Tikunoff, W.J., D.C. Berliner, and R.C. Rist. *Special Study A: An Ethnographic Study of the Forty Classrooms of the BTES Known Sample*. San Francisco: Far West Laboratory for Educational Research and Development, 1975.

White, B.L., J.C. Watts, et al. *Experience and Environment Major Influences in the Development of Young Children (Vol. 1.)*. Englewood Cliffs (N.J.): Prentice-Hall, 1973.

SOME APPROACHES TO INQUIRY IN SCHOOL-COMMUNITY ETHNOGRAPHY

Frederick Erickson
Harvard Graduate School of Education

I have a conviction that the art and the science of qualitative field research can articulate with forms of research generally employed in the study of education. It is important to recognize, however, that when researchers of differing orientations come together to discuss collaboration, there are real differences in premises and methods—in cultures of inquiry—that are likely to make not only for genuine conflict among positions but also for mutual misinterpretation of what the positions of "others" are. In short, one can expect problems of cross-cultural miscommunication and methodological ethnocentrism in symposia of this kind.

Consequently, I begin this paper by considering some of the differences between "qualitative" and "quantitative" types of research. Then I suggest three kinds of strategies for identifying qualitatively derived models and data that could be useful in collaboration across orientations and conclude with some remarks on implications.

An Attempt to Define Some Terms

In emphasizing differences among research orientations it might seem that I am suggesting that the distinction between the *qualitative* and the *quantitative* is the most appropriate way to characterize differences in approach to educational research. I am not sure this is so. There seems to be a distinction among approaches that would be worth making, but it's not clear what that distinction is. The distinction between experimental and naturalistic methods may be a more useful one, but that's not entirely satisfactory either. Fienberg's term "psychostatistician" may become a useful label for those who have a "mainstream" orientation to educational research. However one defines the terms, it seems that the differences among approaches lie not in the presence or absence of quantification *per se* (if one thinks of quantification simply as a means of summarizing information) but in the underlying assumptions of method and proof.

In developing strategies for collaboration across orientations, a fundamental issue is how to get from qualitative study of naturally occurring events in everday life what is essential to such work without making use of it in logically inappropriate ways or so changing the processes of data collection and analysis that the approach we call "qualitative" or "ethnographic" becomes something other than what made it potentially useful and interesting in the first place. This is a concern shared by Hammel who suggests that anthropologists need to learn the languages of other social scientists, notably statistics, not only because quantitative techniques are useful in their own right but also because they enable one to argue in the other people's language, "to point out in no uncertain terms when the assumptions of...mathematical models are violated by the ethnographic facts" (Hammel, 1976:32).

As one who knows next to nothing about quantitative methods—just enough to have had numbers demystified for me—I am impatient with qualitative researchers whose fear of "number crunching" stems from their knowing absolutely nothing about quantitative methods. But some researchers who, like Hammel, are mathematically sophisticated are still concerned that quantitative approaches (and their usual traveling companions, the verification procedures borrowed by social scientists from the physical sciences) may do violence to "ethnographic facts." What are such facts, and what is such violence? What are some good reasons for researchers whose orientation can be labeled "qualitative" to be suspicious of other orientations?

I think that what is essential to qualitative or naturalistic research is not that it avoids the use of frequency data, but that its primary concern is with deciding what makes sense to count—with definitions of the quality of the things of social life. The reluctance of many qualitatively oriented researchers to count things may be related to a theoretically based reluctance to follow Durkheim's injunction (1895:1) to consider social facts as things. Researchers of the Malinowskian tradition in anthropology (and "fieldwork sociologists," "symbolic interactionists," and most recently "ethnomethodologists" in sociology) have been concerned with social *fact* as social *action;* with social *meaning* as residing in and constituted by people's *doing* in everyday life. These meanings are most often discovered through fieldwork by hanging around and watching people carefully and asking them why they do what they do, sometimes asking them as they are in the midst of their doing. Because of this orientation toward social meaning as embedded in the concrete; particular doings of people—doings that include people's intentions and points of view—qualitative researchers are reluctant to see attributes of the doing abstracted from the scene of social action and counted out of context.

I agree with Fienberg in this collection: if qualitative researchers really want to do science and address the problem of the generalizability of insights derived from fieldwork (and if they also have the less disinterested aim of surviving in the arena of policy research), they must become able and willing to count social facts, too. The trick lies in defining carefully what the "facts" are in ways that are precise, reliable, and capable of quantitative summary, yet articulate with the meanings the facts have to the people engaged in everyday life.

The "classical" way qualitative researchers state the social meaning of social facts is through descriptions whose terms have functional relevance within a model of system process. These are descriptions grounded in some theory of the event being described; no such descriptions are *mere* descriptions.

There are many ways to describe what happens in a social event other than in functionally relevant terms. We could, for example, describe the playing of chess in terms of movement in millemeters forward, backward, and sideways on a plane. The behavior of chess pieces on the board could be coded by observers this way with high inter-coder reliability, and the resulting data could be manipulated statistically. Yet by itself, this would tell us nothing about what was going on in the game of chess. We need descriptive categories with functional relevance for the game–checkmate, defense–terms for the *qualities of things* (in an etymologically literal sense) for the *kinds of kinds of things* that are meaningful for an understanding (a working theory) of the game as a whole (cf. Wittgenstein, 1953). To use Eisner's metaphor of *connoisseurship,* no connoisseur would describe chess in functionally irrelevant terms. It should be noted here that by "function" I do not mean it as the term is used by sociologists and anthropologists of various "functionalist" theoretical orientations. Rather, I mean "function" in the sense meant by linguists. This point will become clearer in the subsequent discussion.

Because the statement of functional relevance considers relations between parts and the whole, such work involves systems thinking. It is in this sense that ethnographic work is "holistic," not because of the size of the social unit but because units of analysis are considered analytically as wholes, whether that whole be a community, a school system and its political relations with its various "publics," the relations among those in a school building, or the beginning of one lesson in a single classroom.

Each of these wholes can be considered as a game. Qualitative research seeks to tell us what the game is: what attributes of "things" in the game are functionally relevant to playing the game, what appropriate relations among things there are in the game, and what the game related purposes of the players are. This may seem to researchers trained in other ways to be a claim to omniscience, but there do exist conventional rules of evidence and verification in qualitative analysis. On the basis of our definitions of the quality of things in functionally relevant terms, we can make predictions of how the game will unfold. The test of validity of the qualitatively "grounded" theory of the game is its predictive power; given a finite set of circumstances the theory can tell us what the players could appropriately do next.

I think part of the anxiety of qualitatively oriented researchers about quantification stems from the fear that what will be counted will be functionally irrelevant attributes of the things people are attending to in everyday life. Past history of inter-ethnic conflict among the social sciences may make such anxiety understandable, especially when qualitatively derived models are met by researchers from different orientations shooting from the hip with such questions as, "Where's the evidence?" or "Why are there no verification procedures?" or "What's the sample?" without any reference to the qualitative researcher's main question, which is something like, "What's the game, and how can it be described."

Qualitative researchers might respond by saying, "But my description works–it has predictive validity." This answer overlooks the fact that other researchers were not there in the field to see how events actually did unfold as the researcher finally learned through field experience–a socialization experience–to expect they would. It also ignores the deep distrust or ordinary, unmediated sense impressions that is an epistemological underpinning of standard scientific procedures of verification.

There are genuine differences across research orientations, but they may not be antithetical. One approach to articulation involves considering a distinction analogous to that between functionally relevant and irrelevant terms for description–the distinction between the "emic" and the "etic."

As a way of defining the difference between the etic and the emic, we can consider a difference between kinds of variation among phenomena that can be summarized quantitatively: continuous variation (height, rate of heartbeat) and discontinuous or categorical variation (being tall, medium or short, left and right, presence and absence). In social life, people often treat continuous variation as if it were categorial, chopping up continua into meaningful chunks as if there were discontinuous thresholds–cutting points–along the continua. These perceived thresholds are meaningful in that people in everyday life take action with regard to them so habitually that the actions (and meanings) are conventional.

In everyday interaction, for example, people may treat the phenomenally continuous variable of *height* as if it were discontinuous, categorizing people as short, average, and tall in stature. Units of *stature,* then, would be social facts, defined in terms of people's discriminations of thresholds and the actions they take toward each other on the basis of those discrimations. The continuous variable, *height,* could be measured formally by an arbitrarily defined unit such as the inch or millimeter, capable of reliable use by observers in making low inference judgements. These units of description could be used in valid and reliable ways within a system of technical categorization independent from functional categories or discontinuous "chunks" used by people in thinking of stature.

The distinction between height and stature is analogous to the distinction linguists make between the "etic" and the "emic"–between phenomena considered from the point of view of standardized measurement of form (or if not in terms of measurement, at least in terms of systematic ways in which scientists as external observers define units) and phenomena considered from the functional point of view of the ordinary actor in everyday life (Sapir, 1925; Pike, 1967:35-72; Pelto, 1970:67-87; and the discussion of Sapir's principle of *contrastive relevance* in the comments by Hymes in this volume).

Modern anthropology, sociology, and linguistics have shown great variation among human groups in the emic discrimination and emic salience of physical and social phenomena. Researchers in these disciplines can state systematically what is emic in everyday events and how people take action with regard to the emic. From my point of view, this is what is qualitative about research–statements of the quality of things and relations, descriptions of events in functional terms. Unfortunately, the "literary" narrative form of reporting traditional qualitative research sometimes obscures systematic statements about emic relations. And there is a difference between the particular procedures of discovery and verification employed in deriving and validating such statements of *quality* and those used by other social scientists. The two approaches to verification can articulate, it seems to me, at the point of correspondence between things considered in terms of their form and in terms of their function–the point of correspondence between the etic and the emic. Some aspects of the emic–of thresholds and "chunking" of experience for social use–can be operationally defined and measured etically, technically, in ways that permit low levels of inference in observer judgements.

One can do this for a piano. The intervals of pitch between keys can be specified etically in terms of cycles of vibration per second, like the etic measurements of distance in the example from chess. Such etic measurement of sound is not useful for playing the piano or for analyzing *as a game* the playing of pianos. But if we want to know if two pianos on which piano games are being played are comparable in tuning (so we can state some formal correspondence between the two games), then it is useful and appropriate. While etic measurement cannot tell us what the game *is* (which is a problem of emic, qualitative analysis), it can establish that features of two games are similar in form or different. If one piano were tuned in half steps and the other in quarter tones, we would show evidence in clearly defined operational terms that the game played on the quarter tone piano was not part of the Western European cultural system of music.

I think this working back and forth between etic and emic units of analysis can also be done in studying social events and social knowledge. Some key elements or features can be described etically and become grist for the mills of social scientists from different orientations. (Granted, some key features can't be described etically without doing violence to the uniqueness and spontaneity of everyday life.) This does not relegate qualitative research to initial states in scientific inquiry–to the primitive phase of "exploratory," "intuitive" work. Qualitative researchers have their own procedures for proof, for testing the predictive power of their "working theoretical" models, which can be used to judge the adequacy of qualitative work within the community of discourse and culture of inquiry of qualitative researchers. Work defined as adequate within this community can be used by the community to frame theory at more general levels. At the same time, researchers from different traditions of inquiry can make use of qualitatively derived insights. By learning more about how such insights are derived and stated, those researchers would be in a better position to judge the usefulness of qualitative data and methods for their own work.

For it is not as if researchers whose orientation is not primarily "qualitative" have a malevolent and perverted desire to quantify and derive theory from social "facts" defined in ways that are functionally irrelevant to actors in social life. At that level of generality, I think, the aims of social scientists are similar (with the possible exception of radical behaviorists). What "qualitative" researchers have to offer others is potentially valid insight into functionally relevant definitions of social facts. What "quantitative" researchers have to offer the "qualitatives" is ways of determining the generalizability of qualitative insights, ways of escaping from that tyranny of the single case which Fienberg discusses in this collection.

In the next section, I will consider the first of three strategies discussed in this paper by which information derived from qualitative research can be made useful to other social scientists.

Textual Analysis of Ethnographic Reports

In his introduction to *Argonauts of the Western Pacific,* Malinowski called on ethnographers to report three kinds of descriptive information: (1) an outline of the social anatomy, (2) "imponderabilia of actual life and everyday behavior," and (3) members' points of view, especially as determined from a collection of typical narratives, utterances, folklore, and magical formulae, "as a *corpus inscriptionum,* as documents of native mentality" (1922:22,24).

I think what qualitative research does best and most essentially is to describe key incidents in functionally relevant descriptive terms and place them in some relations to the wider social context, using the key incident as a concrete instance of the workings of abstract principles of social organization.

It is from Malinowski's middle level of "imponderabilia" that the key incidents are derived, usually from field notes. In the research report the generic features of these incidents are highlighted with as much concrete detail as is necessary to make a statement of the relation of the instance to the pattern of the whole. The qualitative researcher's ability to pull out from field notes a key incident, link it to other incidents, phenomena, and theoretical constructs, and write it up so that others can see the generic in the particular, the universal in the concrete, the relation between part and whole (or at least between part and some level of context) may be the most important thing he does. Such selection, description, and interpretation is very emic—indeed, ontological. It involves massive leaps of inference over many different kinds of data from different sources—field notes, documents, elicited texts, demographic information, unstructured interviews, and very possibly survey data. This is a decision process analogous to that of the historian or biographer deciding which incidents among many in a person's life to describe.

Classic examples are Whyte's description of the bowling matches between the corner gang and another young men's club in *Street Corner Society* (1955:318-320) and Malinowski's (1922) description of various incidents involved in the Kula trade network across a number of Melanesian islands. A recent example is Ogbu's story of how rumors spread in a multi-racial, multi-ethnic school community (Ogbu, 1974:133-170). Such incidents of great "working-theoretical" salience may tie together the whole qualitative account. This is so for Whyte's bowling match. As Whyte described it, social relations were played out within and between groups in forms that he claims to have seen repeated in the neighborhood in widely differing group contexts. By describing the bowling match in rich detail, Whyte helps us to see how this instance is generic and how it related to many others.

I have described this process of selective reporting in producing a whole ethnography as *caricature* (Erickson, 1973): an abstraction from the diversity of phenomena as experienced so as to emphasize some features and deemphasize others. To say an image is a caricature is not to deny its validity. Indeed, caricatures in the graphic arts and in literary description can be "truer" than the "actual" life the caricaturist attempts to represent, as all art is in one way or another "truer"—more coherently organized—than life. However, the caricature's validity is of a different epistemological order from that of standard science. It would be fruitless to look for empirical evidence in the phenomenal world for the shape of Richard Nixon's nose as portrayed by some political cartoonists. There may be a "truth" in such portrayal, but it is not amenable to empirical investigation. Still, despite the tendency of ethnographic caricatures to go beyond the bounds of the empirical, the insights they report can have uses for researchers of other orientations.

By what means could these insights be codified and summarized without doing violence to their uniqueness? This is an issue of cross-case comparison in which the unit of analysis is the qualitative research report. The strategy that comes immediately to mind—the approach toward coding "ethnographic facts" across case studies taken by the Human Relations Area Files—may not be the most appropriate.[1]

The qualitative case study is a literary form potentially amenable to some kinds of "text criticism." Perhaps panels of readers with differing points of view—practitioners, policy planners, "quantitatives," and "qualitatives"—could go through a set of case studies and abstract from them key incidents and the interpretaions of those incidents that together constitute the author's "working theoretical" models of social organization in the setting. While working models would vary both in the scale and complexity of the phenomena they attempt to account for, and in orientations of substantive, qualitatively grounded theory (and perhaps general theory) out of which they were constructed, such a review process could suggest common dynamics in operation across individual classrooms, schools, and school communities. It could point to fruitful directions for further research using strategies other than qualitative and might lead to research on aspects of educational processes that had not been considered before. But in addition, such foraging operations over the qualitatively defined "field" phenomenon *education in community life* might simply provide qualitative insight of a kind directly useful for varying audiences concerned with the study and practice of education—new ways of thinking about educational practice, everyday life in schools and communities, everyday aspects of attempts at change.

This is a process of cross case comparison by panel review somewhat similar to that used by Tikunoff, Berliner, and Rist (1975) to identify potentially useful classroom interaction variables from the field notes of participant observers. The results of their subsequent research using reviewer defined variables suggest that review strategies may be useful in articulating the work of qualitatively oriented researchers with that of researchers with differing orientations.

The approach I am suggesting differs from that of Tikunoff, et al., in two ways. First, it differs by entering the process of qualitative, emic definition and modeling at a later stage, after the qualitative researcher has produced a final report by a process of inference and emphasis that gives the report a characteristic shape—the coherence of a caricature. (The decision processes that produce such shape are, I think, intrinsic and essential to the "classic" way of doing qualitative research.) Second, this approach differs by identifying as units of analysis not simply variables but models—ways of thinking about things and their relations that we might for some purposes choose to call *variable sets* and hypotheses, amenable to etic operationalization and testing, but which we might for other purposes choose to consider more loosely and less formally.

Since there seems to be a paucity of new ideas about how to think about what happens in educational, sociocultural, and political processes among actual children, teachers, administrators, and parents in school communities, this approach to finding out what qualitative researchers have to tell us may have considerable merit.

Focused Strategies of Primary Data Collection—An Apologia

The two remaining strategies to be discussed here for collaboration across research orientations involve more focused approaches to qualitative data collection. They may be more compatible with the methods of "psychostatistical" researchers than the textual analysis of qualitative research reports prepared by the "classic" hypertypical lone researcher, working for the most part through informally systematized methods of data collection. Before detailing these two strategies, an explanation is in order.

The decision to use more focused approaches changes the field experience of the researcher. It requires that fieldwork be conceived as a process of actively and consciously directed inquiry in which decisions about researchable problems and the statement of researchable questions are made while the researcher is in the field, rather than at some time after having left it.[2] Specification of data collection strategies while in the field presupposes a conscious theoretical orientation by the researcher—a conscious awareness of one's commitment to points of view derived from substantive theory in social science and from personal theory. Focused data collection also requires knowing something about the setting one is studying through information gathered before entering the setting as well as from first hand experience. This point is made very strongly by Hammel (1976) who, in speaking to anthropologists in particular, says that in the study of complex modern societies it is not useful as a research stategy to pretend to know nothing in advance about the setting one is studying.

Focused data collection strategies are incompatible with the "hypertypical" view of the field research process—in which one begins atheoretically with no prior conceptions about the setting, then "hangs around" letting the setting "tell you what's going on," and finally decides what the problems were after returning from the field. Systematic strategies would seem to leave too little room for intuition and happenstance, for the unmediated richness of field experience. Certainly, there is a danger that focused data collection can freeze the research process prematurely. But greater danger lies in adopting the hypertypical view of field research as highly spontaneous, for I think this view is based on a wrong-headed interpretation of what actually happens in the field. No setting, I would argue, "tells" anybody anything; no questions are generated directly from experience—there are no pure inductions. Research questions come from interaction between experience and some kind of theory, substantive or personal. It is extremely important that qualitative researchers make that interaction as explicit as possible both to their audience in reporting and to themselves while in the field. In no other way can qualitative researchers cumulate knowledge, and in no other way can they avoid a "credibility gap" with other social scientists (cf. Pelto, 1970:1-46).

In short, I am arguing that research on schools can be both qualitative and systematic. We have theory in sociology and anthropology relevant to what happens in American schools. We know a lot already about what happens, and there is no need to pretend methodologically that we don't know anything. On the basis of both kinds of knowledge—the theoretically derived and the experientially derived—we can identify phenomena of emic salience to persons in the setting and operationally define those phenomena in etic terms for systematic procedures of data collection and analysis.

Strategies for doing this can be thought of in two main streams of approach to focused data collection. The first involves working with definitions of what is relevant taken from the existing conscious awareness of school practitioners and from existing literature in social science and educational research. Following Hymes (1976), I will call this approach "ethnographic monitoring." The other approach involves discovering new phenomena of functional relevance—new variables and relationships among variables that may not be accounted for in the conscious awareness of school practitioners but may be suggested by recent research and theory development in the social sciences.

Ethnographic Monitoring

To monitor anything one needs (1) a working model of the whole system or subsystem that is to be monitored and (2) some means of measuring functionally critical features of system process. For example, in constructing a cybernetic system to monitor a home heating plant, one needs a model that specifies some relations between the fuel consumed in the furnace firebox on the one hand and room air temperature on the other and a way to measure amounts of fuel and room air temperature. In a jury-rigged man-machine cybernetic system like the old-fashioned home heating plant, tolerances are fairly wide and the measurement operations can be very informal, approximate, and "emic." One waits until the air temperature is "too cold," then goes to the basement and shovels "enough" coal into the furnace. Not only is precise measurement capability unnecessary but a general theory of system dynamics is not necessary either; one doesn't need a fully developed theory of heat and heat transfer or of combustion to make the system work.

Learning environments, as social systems, are such jury-rigged operations, capable of useful monitoring in fairly crude ways so long as what is measured is functionally critical to the system. The relatively loose ebb and flow of everyday operations in classrooms and other learning environments, together with adaptive learning strategies of children and adults in such settings, results in adaptive knowledge of system processes by members of the setting. They can tell the researcher some of the phenomena that are relevant to monitoring system processes.

In sum, relevant phenomena can be identified on the basis of (1) prior qualitative research in the same or similar settings, (2) the concerns of participants in the setting, or (3) a combination of (1) and (2). These methods of identification resemble those advocated for "formative" evaluation of educational settings (cf. House, 1973; Provus, 1971). Such a focus on issues of value to those affected by and involved in the research makes possible humane and genuinely collaborative relationships between the researcher and the practitioner (cf. Hymes, 1976). After relevant phenomena are identified, they can be "monitored" through systematic, focused observation that generates data capable of quantitative summary.

An example of ethnographic monitoring is the work of Shultz and Harkness (Shultz, n.d.). They were interested in the social contexts in which children spoke Spanish or English in bilingual education programs. This was a salient issue for the program's administrators who were concerned that despite their aim of maintaining children's use of the Spanish language spoken at home, tests showed that the longer the children were in the program the less well they used Spanish. The administrators and researchers wanted to see what "formative" aspects of the children's experience in the bilingual program might be producing these "summative" test results. Such an issue is appropriate for investigation by ethnographic monitoring both because of its salience for program personnel and because an existing body of knowledge—the discipline of linguistics—provides a system of description by which relevant phenomena (in this case "which language is being spoken") can be reliably categorized and monitored.

Shultz and Harkness put a cassette tape recorder in a child's backpack, had each Spanish-speaking child in a bilingual classroom wear it for a half hour, and recorded the child's naturally occurring speech. Analysis of the tapes revealed that the children who had been in the program longest spoke English most frequently to other children (frequency measured in percent of time the child was speaking each language in freely initiated conversation with other students). Monitoring the bilingual teachers, the researchers found that while the teachers conducted the formal "content" instruction half in each language, they gave "procedural" instructions almost exclusively in English. English speaking ability thus became a valuable commodity in the social and political economy of the classroom. The children who had been in the program longer became "brokers" between the newcomers and the teacher, translating the teacher's instructions into Spanish for the newcomers and "speaking up for them" to the teacher in English.

Similar dynamics in the politics of speaking were found by Cohen, Bruyk, and Shultz in a Center for Applied Linguistics study of a classroom in another state. Thus, by turning up in two classrooms, a finding from ethnographic monitoring has escaped the tyranny of the single case. At this point, it would be possible to check generalizability across a number of classrooms. Moreover, the potentially key factor in the system, the teacher's giving procedural instructions entirely in English, can probably be controlled, making possible experimental manipulation of the variable. Or more simply, some teachers might decide to give procedural instructions in Spanish, try to change their behavior, and monitor themselves by wearing the backpack occasionally. They then would control the techniques of the research process and the information generated by it.

There are a number of other aspects of social relations in learning environments that can be operationalized and monitored in quite straightforward ways.

1. **The amounts of time children spend attending**

to the teacher. Aspects of listening behavior that culture members interpret as showing attention—features such as eye contact and postural orientation and stability—can be monitored reliably in detailed coding from videotapes or crude but reliable coding *in situ* by classroom observers. Issues of who attends, how much, and the relation of children's attention behavior to teachers' perceptions of their intelligence and motivation can be addressed, as can culturally patterned differences in ways of showing attention, interest, and comprehension.

2. **The topical relevance of children's discourse with the teacher and with other children**. This variable, mentioned in Cooley's discussion of research on effective teaching, involves both the topical relevance of children's classroom talk and teachers' strategies of fostering topical relevance. Audiotape probably would be necessary and, given the complex ways the meanings of ordinary talk are embedded in the social situation of the moment, videotape recording might be desirable.
3. **Teacher assessment of the intellectual competence of children on the basis of social performance**. This recalls Cazden's "mini-tests" by which teachers informally size up the children, Rist's approach to the same point, and Leacock's "we-they" dichotomy in teachers' folk taxonomies of children. At issue are: (1) the cues teachers employ to make judgments of competence—e.g., how children talk, listen, sit, respond to procedural instructions (and, following Rist, how they smell); (2) the relative differentiation of the teacher's typology of children in the class—the range of "taxons" or dimensions of contrasts in the teacher's cognitive map of the kind of students in the class; and (3) the relative stability of the teacher's typology over time. While monitoring procedures for this topic are not as well developed as for the previous two, and while the judgments required of researchers are more complex, recent literature suggests that the topic is important. Detailed observational records, derived from participant observation alone or in combination with videotaping, as well as interview data would be necessary.
4. **The regularity of classroom activity rhythms**. This can be monitored by timing the speech and body motion of students and teachers and the sustainment of postural configurations by classroom groups. Irregular rhythms of speech and body motion and aperiodic sequences of group postural configurations seem to have social significance. Such occurrences were judged "uncomfortable moments" with high inter-rater reliability in studies of dyadic encounters (Erickson, 1976; Shultz, 1974), and we have found analogous patterns in recent research in classrooms. Ruiz, in a study of videotapes of 100 Head Start classrooms, found that the variable that discriminated best between inexperienced and experienced teachers was the periodicity of the teacher's movement around the room—the duration of each "passage" from one child or group to the next and the duration of time spent with the child or group at a "destination." For experienced teachers there was little variation in the durations of "passages" and "destinations," while for inexperienced teachers there was great variation. We have found similar patterns of temporal regularity among experienced teachers in classrooms in Boston suburbs and in an Indian reservation school in northern Ontario.

Studies Deriving from a Cognitive Theory of Culture and Social Competence

Most of the studies to be reviewed here are recent. They approach issues of the sociocultural organization of learning environments—some of which are new to educational practitioners and researchers and others of which have been addressed before in thinking about education, but in different ways.

The studies share an over-all theoretical frame of reference that is emerging from theoretical and empirical work in anthropology, sociology, linguistics, and social psychology. One way this general approach has been articulated in anthropology is in a cognitive theory of culture. Goodenough has defined culture ideationally as "a system of standards for perceiving, believing, evaluating, and acting" (1971:41). What one has to know in order to act appropriately as a member of a given group includes knowing not only what to do oneself but also how to anticipate the actions of others.[3]

Related definitions can be found in linguistics and sociology. Hymes (1974) notes that knowing how to speak appropriately involves much more than *linguistic competence*, which is Chomsky's (1965) term for a speaker's capacity to employ the sound system and grammar of a language in generating sentences. For Hymes, *linguistic competence* must necessarily entail *social competence*, since acceptable speaking requires the ability to produce not only grammatically appropriate speech but also situationally appropriate speech. Less closely related, but still comparable, is the emphasis on "member's work"—the exercise of practical reasoning in everyday social life—found in the emerging field of sociology called ethnomethodology (Garfinkel, 1967). All these theoretical positions have in common an emphasis on what people need to know in order to do what they do in ordinary social interaction.

They emphasize not simply behavior but the knowledge necessary to produce the behavior.

Educational settings, in schools and families and communities, are especially appropriate for study from this theoretical perspective because they aim to transmit knowledge about how to perceive, believe, evaluate, and act. This transmission takes place largely through the medium of face-to-face interaction (Gearing and Sangree, in press). School classrooms are settings in which special attention is paid to appropriate ways of behaving. Appropriate behavior may be explicity encouraged, an inappropriate behavior—or the absence of appropriate behavior—may be explicitly pointed out and negatively sanctioned. A general question for classroom interaction research is, "What do teachers and children have to know in order to do what they are doing?"

There are two main ways by which we can study people's cultural knowledge—by asking them and by watching them. First I will describe approaches based primarily on asking, then approaches based primarily on watching.

Questionnaires are one way to elicit people's cultural knowledge. In a recent attempt by Jacob and Sanday (1976), questionnaire items were constructed on the basis of Goodenough's general theory of culture and designed to elicit expectations for appropriate school behavior. The instrument was administered to 266 Puerto Rican students and dropouts and 15 teachers in New York, Philadelphia, and Vineland, New Jersey. Interestingly, Jacob and Sanday found through simple statistical analysis that the categories *low hooky-high hooky* discrimated student responses better than the categories *dropout-stayin*, i.e., the responses of some dropouts and stayins were very similar. By moving to this more differentiated classification, they found that the responses of high hooky dropouts and stayins were more similar to those of teachers than were the responses of low hooky students. Low hooky students saw fewer behaviors as acceptable, relative both to teachers and high hooky students. Their low risk strategy of showing up for school is consonant with this view of what is expected of them. One would expect them to adopt low risk strategies in everyday life in the classroom as well.

While the authors acknowledge a number of technical problems involving possible sample and instrument bias, the study is interesting because it reveals a potentially salient dimension of analysis that was not intuitively obvious when the research was begun. Such results can inform further fieldwork, e.g., to see if low hooky students do indeed adopt low risk strategies in the classroom, if high hooky students use their knowledge of the classroom game to make themselves highly visible, and if low versus high hooky dropouts report different kinds of reasons for leaving school. These insights could then be combined with a redesigned questionnaire for further investigation. In such ways, focused primary data collection can inform the researcher during the course of fieldwork, and participant observation can inform the design of focused data collection.

Another approach to eliciting similar information is that of Spradley and his students (Spradley and McCurdey, 1972). They combined field observation with format interviews to elicit the "ethnosemantic" judgments of students about what kinds of activities and social roles there were in classrooms and at school recess. From the interviews, the authors constructed models of emic cultural knowledge about social relationships. They state their models of students' and teachers' "cognitive maps" of rules and plans for everyday school interaction in formal ways whose clear specification of variables could form the basis for further research. While there are serious problems with the use of ethnosemantic elicitation techniques apart from fieldwork, the work of Spradley's undergraduate students is compellingly attractive and its theoretical orientation is clearly articulated. I think this represents a useful approach for fieldworkers and has great potential for researchers from other orientations as well.

I turn now to approaches based primarily on watching—to inferring people's social competence from their social performance. Another implication of the general theoretical position of Goodenough and the others referred to above is that socialization is not simply a matter of reinforcement. The theory assumes that children and adults are actively engaged in constructing emic models of the social worlds in which they find themselves. Especially among ethnomethodologists, an assumption is that socialization is a never ending process, that as people of any age interact they are continuously engaged in telling each other, nonverbally as well as verbally, what is going on. Thus from the study of social behavior (performance) one can infer the social knowledge (competence) necessary to produce the behavior, just as a connoisseur of the symphony orchestra can rigorously and objectively infer the musical knowledge necessary to write a symphony and produce a performance of it. This premise suggests observational methods as a means of primary data collection.

One example of such work is that of Philips (1972, 1975) in which the key analytic notion is *participation structure*, the characteristic "games" or modes of organization by which children and adults conduct everyday interaction. Philips investigated culturally different forms of participation structure through the classic method of participant observation, carefully observing and comparing the interaction of children and adults at home and school on the Warm Springs

Indian reservation. The theoretical orientation was that of Goodenough, Hymes, and Goffman (1964, 1974). Two salient aspects of this work are (1) the comparison of customary participation structures outside school with those inside school and (2) the fully "interactional" character of the analytic model, i.e., the model accounts for what all parties to an interactional event are doing–what one person does while others do what they do.

Philips identified a range of characteristic ways that rights and obligations governing speaking and turn taking were organized and showed cultural differences between participation structures most commonly occuring inside and outside school. A major difference involved the role of the adult or other leader. At school, the leader (the teachers, who were always white) attempted to control all activity, communicative and otherwise, functioning as a switchboard operator, to whom much talk was addressed and by whom all allocation of legitimate turns at speaking was granted. In such a participation structure, the Indian students performed much more situationally inappropriate behavior than did white students in the classroom. For Indian students and adults outside the classroom, Philips reports that participation structures in which one person controls all activity did not occur: "The notion of a single individual being structurally set apart from all others, in anything other than an observer role, and yet still a part of the group organization, is one that [Indian] children probably encounter for the first time in school" (Philips, 1972:391).

Such propositions and working theoretical models are stated in a form entirely appropriate for further focused investigation. They are etic statements of the emic organization of everyday activity. Those who do field research in educational settings can benefit from attempting to state their models as clearly as does Philips.

Currently, Gerald Mohatt and I are using Philip's notion of participation structure to organize our study of the interaction of children and adults on an Odawa reservation in northern Ontario. Using a portable radio microphone and a minimum of visual "camera editing" we have been making continuous videotapes of interaction at home and in two school classrooms. All the school children are Indian; one of the teachers is Indian, the other white. We are interested in the extent to which (1) the white teacher organizes participation structures involving all or some children in ways similar to Philips's models of teacher-student interaction and (2) the Indian teacher organizes participation structures differently from Philips's models. Philips's work would lead one to expect that the Indian teacher might organize participation structures without putting herself constantly in a position of absolute control over all activity, and on the basis of preliminary analysis of our tapes, that seems indeed to be the case. Moreover, our tapes of family interaction at home show participation structures Philips found characteristic of interaction outside shcool at Warm Springs.

By direct analysis of minimally edited behavior records (audiotapes, videotapes, cinema film), models of the social organization of interaction can be tested carefully for validity and generalizability. Systematic sampling of recurrent events in the daily rounds of teachers and students is possible, and the data generated can be operationally defined in etic terms for which high inter-rater reliability can be demonstrated (cf. Erickson, 1976a). Such data is amenable both to carefully controlled logical analysis as done by linguists and to quantitative summary and analysis.

Interaction analysis directly from behavior records enables the researcher to observe repeatedly each "strip" of interaction being investigated. This can prevent premature "typification" in constructing models. One can stay in touch with discrepant cases that do not quite fit an initially undifferentiated analytic model, adjusting the model to take account of variation that is not trivial by stating "variable rules" and "exceptions to the rules," as well as more general patterns. The work of Mehan, et al., (1976) is exemplary in this regard. In their analysis of instructional sequencing in classroom lessons, they are able to account for systematic variation in their data, accounting for every case in their sample by methods of discrepant case analysis.

While Mehan, et. al. have focused primarily on verbal interaction, McDermott (in press) has investigated nonverbal interaction in a comparative study of the social organization of taking turns at reading in "high" and "low" reading groups in a first grade classroom. In a related approach Gumperz has studied "contextualization-cueing"–the verbal and nonverbal cues by which people signal each other how to interpret what they are saying as they say it (Gumperz and Herasimchuk, 1972; Gumperz, 1976).[4]

A final example of current research in these directions can be characterized as eclectic in the extreme. Shultz, Florio, Walsh, Bremme, and I have been investigating the participation structures and social competence of one kindergarten-first grade teacher and her students (Shultz, 1976). Over a two-year period, we have videotaped in the classroom for a total of 72 hours of tape and, to a much lesser extent, in children's homes.

A relationship of close collaboration with Walsh, the teacher, has evolved (Florio and Walsh, 1976), enabling us to integrate the humane relationship of dialogue with a key informant that has been essential to much ethnographic research with the more systematic and "distanced" observational methods employed in the direct analysis of behavior records.

We also have interviewed the teacher in spontaneous conversation, in formal interviews, and in "viewing sessions" in which we watch and discuss videotaped excerpts of classroom happenings identified as potential "key incidents." Shultz, Bremme, and Florio have analyzed the videotapes, preparing intensive case studies of verbal and nonverbal behavior in key incidents that highlight fine details of participation structure: how children get turns at speaking, how the mutual rights and obligations of those engaged in interaction shift from moment to moment (cf. Cicourel 1972), and how brief and transient "subcontexts" are played out within larger strips of activity—moments in which what was socially appropriate the moment before is no longer appropriate.

To get on in school, teachers and children need a social "radar" for monitoring the culturally patterned contextualization cues that signal subtle shifts in context of situation from moment to moment. Researchers studying the role of communicative competence in classroom interaction need methods for the empirical investigation of such contextual shifting. Because of the relative indeterminacy of segment boundaries or "junctures" between emically salient "chunks" of everyday interaction, the empirical study of contextual shifting is a problem deserving continuing basic research. School classrooms are highly appropriate settings for such research.

Case studies of classroom participation structure derived from direct analysis of behavior records are a means of producing etic data which can be quantitatively summarized yet which also can be articulated with categories of emic structures relevant to the point of view and purposes of teacher and students. Units of data and combinations of units that are identified through videotape analysis and operationally defined in etic terms can be tested in "viewing sessions" for congruence with the teacher's ways of talking about the events. Thus for a given classroom event, points of formal correspondence can be shown between (1) the teacher's emic model of the event, as elicited in interviews, (2) the reseracher's "emic/etic" model of the event identified from direct analysis of the behavior record, and (3) etically defined measurement operations that produce frequency data. (Note the correspondence of this approach with Bartlett's ladder diagram for the process of scientific inquiry found in Fienberg's paper in this collection.)

All of the studies reviewed in this section have addressed the relationship of social or communicative competence to the enactment of everyday life in classrooms.[5] The theoretical and methodological orientations of most of these studies allow the researcher to stay in touch not only with the concrete details of the enactment of social life, and with the "rules" for enactment that are usually studied and reported by social scientists—the customary patterns or normative order according to which social scenes are played out day after day—but also with the creativity and spontaneity involved in recurrent performance by which the old and familiar is continually made new and chosen. Using a variety of methods, most of which permit quantitative summary of data, these reserachers are attempting to discover new qualitative knowledge—new aspects of what children need to know in doing going to school and of what teachers need to know in doing teaching.

In concluding, it is appropriate to ask how all this relates to issues in the mainstream of educational research and to issues of quantitative method.

Despite surface differences, there seems to be considerable convergence between the work reviewed in the previous section and the work of Smith and Geoffrey (1968), Smith and Carpenter (1972), and Kounin (1972), on the one hand, and of Barker and Gump and their associates on the other (Barker, 1965, 1968; Gump, 1969; Gump and Ross, 1975; Gump and Good, 1976).

In attempting to conceptualize the process of teaching, Smith has emphasized ringmastership and its components—awareness, pacing, sequential smoothness, and teaching in motion. Kounin has identified similar dimensions of the process of classroom management—momentum, withitness, smoothness, overlapping, and variety. Both these conceptual schemes emphasize the timing of activity and point up what may be one of the most salient ethnographic "facts" about life in classrooms—that there always seems to be more than one thing at a time happening. Effective teachers seem to be able to handle this multiplicity of events. Some students, whether because of differences in culture, temperament, or ability, seem to be able to handle the multiplicity better than others and to perform more effectively, socially and academically. As Rist reports (1970), the social behavior of children in classrooms establishes social identities for them from the point of view of the teacher, and these social identities seem to correlate with academic achievement and form a basis for tracking students in the early grades (see also Mehan, 1975, and Leiter, 1975).

Philips (1972) provides another way of formally describing the social organization of the multiplicity of events in the classroom—a way that permits specification of variation across different structures of participation, different social environments for learning. Those environments may be the key unit of analysis for the study of classroom interaction. This point is made in different terms by Kelly (1969) and by Gump (1969:201), who notes: "The root problem in ecological psychology is conceptualization of the environment. The study of the subject's behavior in his natural habitat is *not* the same as the study of natural habitats." In recent work, Gump has reported ways of characterizing whole environments using quantitative data (Gump and Ross, 1975; Gump and Good, 1976).

We are approaching the time when we can construct comparative typologies or models of whole classroom learning environments and identify styles of classroom management by teachers and classroom behavior by children within the context of the over-all classroom environment. When this becomes possible, we can investigate what styles of being a student "go with" what styles of teaching and how these different forms of social relationship in classrooms correlate with the outcome measures of achievement traditionally measured in educational research. Then we can begin to learn ways of matching kinds of teachers, kinds of children, and kinds of learning environments that result in optimal outcomes.

The other issue I want to address briefly as a postscript concerns methods of quantification. On this subject I have only minimal technical knowledge. But provided one collects primary data so there is some correspondence between the emic ways people have of ordering interaction in everyday life and etic ways of operationalizing variables, it would seem that there is no inherent contradiction in using quantitative methods in qualitative research. I have argued elsewhere (Erickson, 1976) that the statistical techniques appropriate for the analysis of qualitatively derived models and data may well be extremely simple techniques—the chi-square, the Mann-Whitney two-tailed test in the analysis of "categorical" data, and two and three way analysis of variance.[6]

The purpose of such quantitative analysis is simply to demonstrate the validity of one's analytic models—models in which, because of their grounding in qualitative observation, one knows a good deal about the expected variation before conducting statistical analysis. I have arrived at this contention in collaboration with Shultz and find support for it from Pelto, who has suggested that if in addition one wants to use more elaborate statistical techniques in the analysis of qualitatively derived data, the approaches of Bayesian statistics—in which one can specify expected ranges of variation and adjust these expectations during the process of analysis—may be more appropriate than the approaches of classical statistics (personal communication, July 11, 1976).

Of this last point I am not technically competent to judge. It would seem that qualitative researchers could benefit from extended consideration of such issues of technique together with experts in statistics, perhaps in summer institutes or working conferences in which there would be adequate time to learn more about each other's expertise. That dialogue is long overdue.

Notes

1. I am indebted for this point and for the discussion that follows to Robert Herriot, personal communication, July 1, 1976.
2. Several recent writers on qualitative methods who have emphasized the role of conscious inquiry in fieldwork are Denzin (1970), McCall and Simmons (1969), Pelto (1970), Schatzman and Strauss (1974), and Runcie (1976).
3. For additional expositions of this position see Wallace, 1970:1-45, and the introductory essay in Spradley (1972).
4. The recent work of Kendon (1967), Duncan (1972), and Mayo and La France (1975) also deal with contextualization-cuing processes (under different names), as does my recent work on functions of postural positions and of speech and body motion rhythms in the regulation of interaction in school counseling interviews (Erickson, 1975, 1976a, 1976b).
5. For a review of additional related studies and recommended directions for research, see N. L. Gage (ed.), *NIE Conference on Studies in Teaching: Panel 5*, "Teaching as a Linguistic Process in a Cultural Setting," December, 1974.
6. The work of Duncan (1972) is instructive in this regard. In his analysis of the functions of nonverbal cues in conversational turn taking, he finds chi-square values at the .0001 level of statistical significance.

References

Barker, R.G. *Ecological Psychology*. Stanford: Stanford University Press, 1968.

Barker, R.G. "Explorations in Ecological Psychology." *American Psychologist* 20:1-14, 1965.

Chomsky, N. *Aspects of the Theory of Syntax*. Cambridge: M.I.T. Press, 1965.

Cicourel, A.V. "Basic and Normative Rules in the Negotiation of Status and Role." In D. Sudnow (ed.). *Studies in Social Interaction*. New York: Free Press, 1972.

Denzin, N. *The Research Act: Introduction to Sociological Methods*. Chicago: Aldine, 1970.

Duncan, S. "Some Signals and Rules for Taking Speaking Turns in Conversations." *Journal of Personality and Social Psychology* 23:283-92, 1972.

Durkheim, E. *The Rules of Sociological Method*. Glencoe (Ill.): Free Press, 1958. (Original, 1895).

Erickson, F. "What Makes School Ethnography 'Ethnographic'?" *Council on Anthropology and Education Newsletter* 4(2):10-19, 1973.

Erickson, F. "Gatekeeping Encounters: A Social Selection Process." In P.R. Sanday (ed.). *Anthropology and the Public Interest*. New York: Academic Press, 1976a.

Erickson, F. "One Function of Proxemic Shifts in Face to Face Interaction." In A. Kendon, M.R. Key, and R. Harris (eds.). *The Organization of Behavior in Face to Face Interaction*. The Hague/Chicago: Mouton-Aldine, 1976b.

Florio, S., and M. Walsh. "The Teacher as Colleague in Classroom Research." Paper presented at the Annual

Meeting of the American Educational Research Association, San Francisco, April 21, 1976.

Garfinkel, H. *Studies in Ethnomethodology.* Englewood Cliffs, N.J.: Prentice-Hall, 1967.

Gearing, F. and L. Sangree. *Toward a General Theory of Education.* The Hague/Chicago: Mouton/Aldine, (in press).

Goffman, E. "The Neglected Situation." In J. J. Gumperz and D. Hymes (eds.). "The Ethnography of Communication." *American Anthropologist* 66(2):133-36, 1964.

Goffman, E. *Frame Analysis.* New York: Harper Colophon Books, 1974.

Goodenough, W. *Culture, Language, and Society* (Addison-Wesley Module). Reading (Mass.): Addison-Wesley, 1971.

Gump, P.V. "Intra-Setting Analysis: The Third Grade Classroom as a Special But Instructive Case" In E. P. Willens and H. L. Raush (eds.). *Naturalistic Viewpoints in Psychological Research.* New York: Holt, Rinehart and Winston, 1969.

Gump, P.V., with L.R. Good. "Environments Operating in Open Space and Traditionally Designed Schools." *Journal of Architectural Research* 5(1):20-27, 1976.

Gump, P.V., and R. Ross. "Problems and Possibilities in Measurement of School Environments." Paper presented at Conference of the International Society for the Study of Behavioural Development, University of Surrey, Guildford, England, July, 1975.

Gumperz, J. "Language, Communication, and Public Negotiation." In P.R. Sanday (ed.). *Anthropology and the Public Interest.* New York: Academic Press, 1976.

Gumperz, J., and E. Herasimchuk. "The Conversational Analysis of Social Meaning: A Study of Classroom Interaction." In R. Shuy (ed.). *Sociolinguistics: Current Trends and Prospects* (23rd Annual Round Table, Monograph Series on Languages and Linguistics). Washington: Georgetown University Press, 1970. Also in M. Sanches and B. Blount, *Sociocultural Dimensions of Language Use.* New York: Academic Press, 1975.

House, E.R. *School Evaluation: The Politics and Process.* Berkeley: McCutchan, 1973.

Hymes, D. *Foundations in Sociolinguistics.* Philadelphia: University of Pennsylvania Press, 1974a.

Hymes, D. *Reinventing Anthropology.* New York: Random House (Vintage Books), 1974b.

Hymes, D. "Ethnographic Monitoring." Paper for the Symposium, "Language Development in a Bilingual Setting." Multilingual/Multicultural Materials Development Center, California State Polytechnic University, Pomona, California, March 19-21, 1976.

Jacob, E. and P.R. Sanday. "Dropping Out: A Strategy for Coping with Cultural Pluralism." In P.R. Sanday (ed.). *Anthropology and the Public Interest: Fieldwork and Theory.* New York: Academic Press, 1976.

Kelly, J.G. "Naturalistic Observations in Contrasting Social Environments." In E.P. Willens and H.L. Raush (eds.). *Naturalistic Viewpoints in Psychological Research.* New York: Holt, Rinehart and Winston, 1969.

Kendon, A. "Some Functions of Gaze Direction in Face to Face Interaction." *Acta Psychologica* 26:22-63, 1967.

Kounin, J.S. *Discipline and Group Management in Classrooms.* New York: Holt, Rinehart and Winston, 1970.

Leiter, K. "Ad Hocing in the Schools: A Study of Placement Practices in the Kindergartens of Two Schools." In A. Cicourel, et al. *Language Use and School Performance.* New York: Academic Press, 1975.

McCall, G.J., and J.L. Simmons. *Issues in Participant Observation.* Reading (Mass.): Addison-Wesley, 1969.

McDermott, R. "Relating and Learning: An Analysis of Two Classroom Reading Groups." In Roger Shuy (ed.). *Linguistics and Reading.* Rowley (Mass.): Newbury House, (in process).

Malinowski, B. *Argonauts of the Western Pacific.* New York: Dutton Paperbacks, 1961. (Original ed., 1922.)

Mehan, H. "Accomplishing Classroom Lessons." In A. Cicourel, et al. *Language Use and School Performance.* New York: Academic Press, 1975.

Mehan, H., S. Fisher, and N. Maroules. *The Social Organization of Classroom Lessons.* A Final Report Submitted to the Ford Foundation, 1976.

Ogbu, J. *The Next Generation: An Ethnography of Education in an Urban Neighborhood.* New York: Academic Press, 1974.

Pelto, P. *Anthropological Research: The Structure of Inquiry.* New York: Harcourt, Brace, 1970.

Philips, S.U. "Participant Structures and Communicative Competence: Warm Springs Children in Community and Classroom." In C. Cazden, D. Hymes, and V. John (eds.). *Functions of Language in the Classroom.* New York: Teachers College Press, 1972.

Philips, S. "The Invisible Culture: Communication in Classroom and Community on the Warm Springs Reservation." University of Pennsylvania, 1975 (unpublished dissertation).

Pike, K. *Language in Relation to a Unified Theory of the Structure of Human Behavior.* The Hague: Mouton, 1967.

Provus, M. *Discrepancy Evaluation for Educational Program Improvement and Assessment.* Berkeley: McCutchan, 1971.

Rist, R. "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education." *Harvard Educational Review* 40:411-450, 1970.

Runcie, J.F. *Experiencing Social Research.* Homewood (Ill.): Dorsey, 1976.

Sapir, E. "Sound Patterns in Language." *Language* 1:37-51, 1925.

Schatzman, L., and A. Strauss. *Field Research: Strategies for a Natural Sociology.* Englewood Cliffs, N.J.: Prentice-Hall, 1974.

Shultz, J. "It's Not Whether You Win or Lose, It's How You Play the Game: A Microethnographic Analysis of Game Playing in a Kindergarten/First Grade Classroom." Working paper, Newton Classroom Interaction Project, Harvard Graduate School of Education, 1976.

Shultz, J. "Language Use in Bilingual Classrooms." Harvard Graduate School of Education, (undated).

Smith, L.M., and W. Geoffrey. *The Complexities of an Urban Classroom.* New York: Holt, Rinehart and Winston, 1968.

Smith, L.M., and P.C. Carpenter. *General Reinforcement Package Project: Qualitative Observation and Interpretation.* St. Ann (Mo.): CEMREL, Inc, 1972.

Spradley, J.P., and D.W. McCurdy. *The Cultural Experience: Ethnography in Complex Society.* Chicago: Science Research Associates, 1972.

Spradley, J.P. *Culture and Cognition: Rules, Maps and Plans.* San Francisco: Chandler, 1972

Tikunoff, W., D.C. Berliner, and R.C. Rist. *Special Study A: An Ethnographic Study of the Forty Classrooms of the BTES Known Sample.* San Francisco: Far West Laboratory for Educational Research and Development, 1975.

Wallace, A.F.C. *Culture and Personality* (2nd ed.). New York: Random House, 1970.

Whyte, W.F. *Street Corner Society* (2nd ed.). Chicago: University of Chicago Press, 1955.

Wittgenstein, L. *Philosophical Investigations.* New York: Macmillan, 1953.

CRITIQUE

George W. Fairgrieve
Clayton, Missouri, Public Schools

Basically, I have no quarrel with Stephen Feinberg's paper except that it seems to restrict the search for the independent variables—for the "interactions and characteristics of the classroom and the teaching-learning participants as a causal factor" (to use the terminology of the workshop problem statement).

In stating that the ethnographer chooses to investigate such interactions using the method of direct observation for data collection, Feinberg is essentially correct. And he states it well when he says that what we wish to do, whatever our position in the quantitative/qualitative debate, is "to make proper inferences from data." But the problem becomes thorny when we must deal with Jeffrey's statement, cited by Feinberg: "No matter what the subject-matter, the fundamental principles of the method must be the same. There must be a uniform standard of validity for all hypotheses, irrespective of the subject."

A central issue here is the purpose of the ethnographer. The ethnographer generally has as his goal the development of hypotheses, not their testing, although that testing may be carried out subsequently. For example, in Smith and Goeffrey there is a statement of the relationship between teacher awareness and pupil sentiment. Smith and Klein (1969) took this model and attempted to test it empirically in a study involving 69 teachers and their students. In reporting the study, Smith (1971) underscored the methodological approach:

> We have found the field study important for the generation of concepts, hypotheses, and miniature theories. These ideas can then be operationalized, quantified, and tested in broad-scale correlated analyses as we did with "teacher awareness." Hopefully also, these ideas can be moved into even more rigorous experimental designs. Only after that kind of endeavor can one have confidence that the findings pertain to more than our one case.

Perhaps this process of explicit model building and testing would alleviate some of Feinberg's concerns about the drawing of inferences from observational data and about probabilistic model building and single case studies. I would argue that observational studies are needed and would contend that Feinberg's large-scale randomized, controlled field trials are fine so long as they do not replace observational studies. Indeed, such trials, if they can be undertaken, must be based upon observational studies—or some data, hunches, or whatever—that have been developed elsewhere.

Stephen Feinberg's paper champions a standard quantitative position which has been with us in education a long time. Perhaps not so long with us, and certainly not as well regarded, is the ethnographic methodology, the qualitative side of the controversy, which Frederick Erickson championed.

Erickson argues strongly for the legitimacy of ethnographic methodology. His basic position is that the ethnographer is necessary to determine "what makes sense to count." The process of qualitative research described by Erickson, in which key incidents serve to elucidate the working of abstract principles of social organization, does require massive leaps of inference for many different kinds of data. But where would we be without this process.

Erickson makes an effort to reconcile quantitative and qualitative methodologies. He suggests a process of "text criticism" to verify the ethnographer's insights and goes on to propose a combination of approaches in which "points of formal correspondence" are shown for the teacher's emic model of a classroom event, the researcher's "emic/etic" model, and etically defined measurement operations that produce frequency data.

It would seem that Erickson's strategy might offer hope for resolving the controversy between qualitative and quantitative methodology. However, Smith (1971) proposed a similar model which included: (1) experimental design with pre and post tests of achievement, control groups, and inferential statistics; (2) social survey with interviews and questionnaires and random sampling of program relevant individuals with qualification and cross tabulation of responses; and (3) participant observation study. There is little evidence that Smith's proposal had a major impact upon his fellow educational researchers. Let's hope that Erickson's recommendation receives more attention.

This brings me to my concern as a practitioner. As I read the papers by Fienberg and Erickson, I was reminded of Homan's statement in *The Human Group* (1950) about the issue of clinical vs. analytical science as follows:

> It is high time we knew the difference between clinical and analytical science. Clinical science is what a doctor uses at his patient's bedside. There, the doctor cannot afford to leave out of account anything in the patient's condition that he can see or test... It may be the clue to the complex. Of course the doctor has some general theories at the back of his mind... These doctrines may turn out to be useful, but he cannot, at the outset, let them master his thinking. They may not take into consideration, and so may prevent his noticing, the crucial fact in the case before him.
>
> In action we must always be clinical. An analytical science is for understanding but not for action, at least not directly. It picks out a few of the factors at work in

> particular situations and describes systematically the relations between these factors. Only by cutting down the number of factors considered can it achieve this systematic description. It is general, but it is abstract... When progress is rapid, clinical and analytical science help one another. The clinicians tell the analysts what the latter have left out. The analysts need the most brutal reminders because they are always so charmed with their pictures they mistake them for the real thing. On the other hand, the analysts' generalizations often suggest where the clinicians should look more closely. Both the clinician and the analyst are needed.

While the parallel with the controversy over qualitative vs. quantitative methodology is not complete, I wonder if the present controversy might not be the same old debate under different labels. Certainly, from my viewpoint as a practitioner, I would side with Homans that both views are needed.

Whether one's audience is the school board, community, or staff, two types of questions almost always are asked about the schools. One is concerned with how well students achieve—the product; the other, with what happens to students—the process. Program description and evaluation cannot be confined to the traditional pretest, treatment, posttest model unless the only desired outcome is improved test scores. This is especially true when schools seek to innovate since many questions arise not only about the mechanics of the program and test results but also what happens to the student within a program—about what life is like in the classroom.

As practitioners, we need what both qualitative and quantitative research methodologies can contribute. And I would argue that what we need is important. Most research, after all, is conducted with tax money and focuses on public school students. It must be relevant and usable if researchers are to have funds and populations with which to work. I hope ways can be found for ethnographer and psychostatistician to collaborate rather than compete, for this is what we need to increase our chances for making better decisions and improving education.

References

Homans, G. *The Human Group.* New York: Harcourt, Brace & World, 1950.

Smith, L.M. "Participant Observation and Evaluation Strategies." Paper presented at the Annual Meeting of the American Educational Research Association, New York, February, 1971.

Smith, L.M., and Kleine, P. "Teacher Awareness: Social Cognition in the Classroom." *School Review* 77:245-56, 1969.

CRITIQUE

Elliot W. Eisner
Stanford University

Along with virtually all of the other papers in this symposium dealing with qualitative approaches to research in education, Erickson's paper implies that ethnography is *the* field that exemplifies the use of such methods. A second assumption implied in both Fienberg's and Erickson's papers is that whether qualitative or quantitative approaches are used, scientific forms of knowledge are the desired end. I want to claim that ethnography in no way exhausts the fields that employ qualitative methods and that science is not the only nor necessarily "the best" model for seeking and disclosing our understanding of what goes on in educational settings.

As a paradigm case of qualitative inquiry, consider the work of artists. In whatever field, artists primarily are concerned with creating essentially qualitative forms. They formulate qualitative ends-in-view, some vision of what is desired, and then arrange component qualities to achieve such ends. This process, as a whole, is one of qualitative inquiry.

Another form of qualitative inquiry is found in the work of critics. Criticism is an empirical undertaking in that it reveals not abstractions but qualities and their relationships. Criticism can take anything as its subject matter. I believe that the creation of educational criticism could provide a kind of utility that scientifically oriented studies and quantitative treatment of phenomena neglect.

A necessary condition of useful educational criticism is educational connoisseurship. Generally defined, connoisseurship is the art of appreciation. It is essential to criticism because without the ability to perceive what is subtle and important, criticism is likely to be superficial or even empty. Development of educational connoisseurship requires an ability to perceive the subtle particulars that participate in educational life and to recognize the way those particulars form a part of the structure within the classroom. Erickson makes this point well in his apt discussion of the chess game and the need for "descriptive categories with functional relevance for the game."

Educational criticism has three major aspects—description, interpretation, and evaluation. Although there is no sharp line among them, there is a difference in focus and emphasis. The descriptive aspect aims at the vivid rendering of the qualities perceived in the situation. The interpretive attempts to provide an understanding of what has been rendered by using, among other things, ideas, concepts, models, and theories from the social sciences. The evaluative seeks to assess the educational significance of the events or objects described. The critic's major function here is to apply educational criteria so that judgments are grounded in some view of what counts within an educational perspective.

Let me turn now from the point I've been making—that qualitative inquiry in education is not limited to ethnographic methodology—to the second assumption implied in the papers by Fienberg and Erickson. This is the assumption that scientific forms of knowledge are the desired end of inquiry.

Since the early work of E.D. Thorndike, American educational research has been essentially behavioristic in its psychology and operationalist in its philosophy. To "know" has meant to make statements couched in the form of propositions which can be appraised by logical criteria. But since logic is essentially a tool for determining consistency among propositions, something more is needed if propositions are to be more than merely consistent. If they are to make true statements about the world, referents for those statements must be located in that world. And since in empirical matters, observation is subject to biases of one sort or another, observations had to be operationalized through reliable, quantitative standardized procedures since these were least likely to suffer from unreliability.

For generations, this concept of the meaning of knowledge has dominated educational inquiry. Doctoral programs have socialized students to believe that the only procedures one could use to obtain knowledge are scientific and that respectable inquiry in education, at least empirical inquiry, is scientific in character. To use other methods—to employ metaphor, analogy, simile, or poetic devices—has been to lack rigor. Indeed, Fienberg quotes Jeffreys to make this point: "There must be a uniform standard of validity for all hypotheses, irrespective of the subject. Different laws may hold in different subjects, but they must be tested by the same criteria;...." To put the case more strongly, students have been professionally socialized *not* to consider alternatives to behavioristic positivism as a means for understanding educational practice.

How might we compare qualitative and quantitative modes of inquiry in education? It is patently clear that both attend to qualities emerging within educational settings. For example, the investigator interested in the incidence of teacher approval in the classroom must attend to the qualities of such approval to secure data. Furthermore, both the quantitative and qualitative inquirer will interpret information secured from the classroom and, in general, make some value judgments about its educational meaning (although the qualitative inquirer may be more likely to do this).

The two modes differ, I believe, in two respects. First and most important, they differ in the language of disclosure. The quantitative inquirer is obliged to transform the qualities perceived into quantitative terms so they can be treated with statistical tools. This is evident throughout Stephen Fienberg's paper. The qualitative inquirer, on the other hand, uses a mode of disclosure that allows one to envision and experience what one has not experienced directly. The use of this mode of disclosure is illustrated at several points in Erickson's paper. Thus, what most radically distinguishes the two forms of inquiry is how they choose to inform the world about what they have seen.

The second feature distinguishing quantitative from qualitative inquiry is the tendency of the former to structure procedures and to define in advance what shall be attended to a significantly greater degree than the latter. This distinction is evident in most of the papers presented at this symposium.

In making these differentiations, I am in no way arguing that one approach is superior to the other. One approach *is* superior to the other, but only with respect to the nature of the problem one chooses to investigate. It is in this judgment—the question of when and for what purposes each mode of inquiry is appropriate—that the toughest intellectual task is posed in laying out a strategy for the investigation of educational problems.

In *An Essay on Man*, Ernest Cassirer points out that a scientific perspective without an artistic one, or an artistic perspective without a scientific one, leads to monocular vision; both are necessary to have depth. Cassirer's plea for binocular vision through complementary forms of inquiry is one I would echo. One mode of conception and one form of disclosure is simply inadequate to exhaust the richness of educational life.

ASSESSING LANGUAGE DEVELOPMENT—WRITTEN AND/OR ORAL

Much dissatisfaction has been growing over the sole use of achievement tests for determining the success of written and/or oral language development in children. At the same time, the public continues to clamor for evidence that children are learning, and therefore that teachers are teaching and the schools are functioning as intended. Means must be developed that describe and identify learning outcomes. What are some promising practices that can do this, and how can they be utilized as alternatives to only testing for achievement, particularly in the area of assessing written and/or oral language development?

QUANTITATIVE LANGUAGE DATA: A CASE FOR AND SOME WARNINGS AGAINST

Roger W. Shuy
Georgetown University and
Center for Applied Linguistics

Linguistics is a relative newcomer to the academic world, and for this reason, it has undergone rapid change and continues to be subject to new paradigms and cynosures.[1] As long as the goal of linguistics was to write a grammar of languages and as long as the concept of grammar was focused on abstract generalities, quantitative analysis was never very important to linguists. In their concentrated effort to find universals, linguists tended to ignore particulars. In their attempt to find underlying rules, they tended to overlook interesting patterns on the surface. In their efforts to develop a viable theory, they tended to say that everything else was trivial.

Even in the days before generative grammar, however, there was little concern with quantitative analysis of language. A typical approach to grammar writing was to work with an informant and ask questions for weeks, months, or years, depending on the fieldworker's time and energy and the informant's patience. The occasional large-scale language surveys, such as the Linguistic Atlas of the United States and Canada, used a relatively large number of informants, but usually only one occurrence of a given linguistic feature was studied. For example, one of the Atlas vocabulary questions asks, "What do you fry eggs in?" The expected responses included the lexical items, *skillet, frying pan, spider,* etc. Once an informant responded *skillet,* the topic was dropped even though it is quite possible such a response would be given in only 60% of its possible occurrencs given adequate opportunity for it to occur naturally in non-interview conversation. Other peaks of linguistic interest in quantitative measures can be noted, such as the concern for lexicostatistics, but generally speaking, quantitative studies were not common in the field.

Quantitative Analysis in Language Variability

At least three things began to change this state of affairs in linguistics: (1) the general braodening of interest which began to develop in the sixties, leading to new kinds of interdisciplinary studies; (2) development of interest in problems of minority peoples, especially in the schools; and (3) general discomfort with separating the study of formal grammar from the semantic aspects of language.

Linguists began to take an interest in language in social contexts and in urban language in particular. They began to understand that new data-gathering techniques and new modes of analysis were needed. Meanwhile, linguists who had been interested in language variation as it is found in the creolization and pidginization of language also began to apply their knowledge to urban social dialect, particularly the language of urban, northern Vernacular Black English speakers, often providing important historical backgrounds for language change and analytical insights. The general focus, of course, was on variability, not on abstract uniformity, and the critical measurement point was provided by the variability offered by Vernacular Black English.

Several important characteristics contrast these recent approaches (Labov, 1966; Shuy, Wolfram and Riley, 1967; Wolfram, 1973; Fasold, 1972; and others) from the study of variation carried out by dialect geographers. In addition to a more sophisticated sampling technique, the new social dialect study attempted to provide a less structured and more natural body of data from each informant. The need for large amounts of continuous free conversation was stressed, and the single item response formats of the Atlas questionnaire were downplayed. Deliberate efforts were made to obtain speech samples in different styles (narrative, reading, casual, formal, etc.), and considerable effort was put into precise identification of the informant's socioeconomic status (strategies usually borrowed from sociology).

Dialectologists unfamiliar with these methodologies initially were distressed by what appeared to be a sellout to the sociologists (emphasis on statistics, sampling, etc.) and by an initial confusion about what such strategies implied. For example, the new descriptions of Vernacular Black English included features which mainstream dialectologists knew to be characteristic of whites as well. In some quarters, in fact, it was observed that there really was no difference between the speech of blacks and whites—for example, in the South (Williamson, 1971). If one used a methodology which ignored the frequency of occurrence of given linguistic features, such an observation would be natural. But the newer research in social dialects pointed out that in communities in which a given feature, even a stigmatized feature, was used by more than one SES, race, or group of any social category, a clearly discernible stratification of a quantitative nature often was evident.

Figure 1 clearly demonstrates an instance of such stratification for the use of multiple negation. Note that the occurrence of multiple negation across four SES groups is maintained, but that blacks use multiple negatives at a higher frequency than do whites. Further information reveals that men use them at a higher rate than do women. Such data cannot tell us that blacks use multiple negatives and that whites do not nor that men use them and women do not. But it *does* offer rich information about the tendencies toward higher or lower variability usage than we ever could obtain from a methodology which offered only a single instance of such usage as evidence of its use or nonuse.[2]

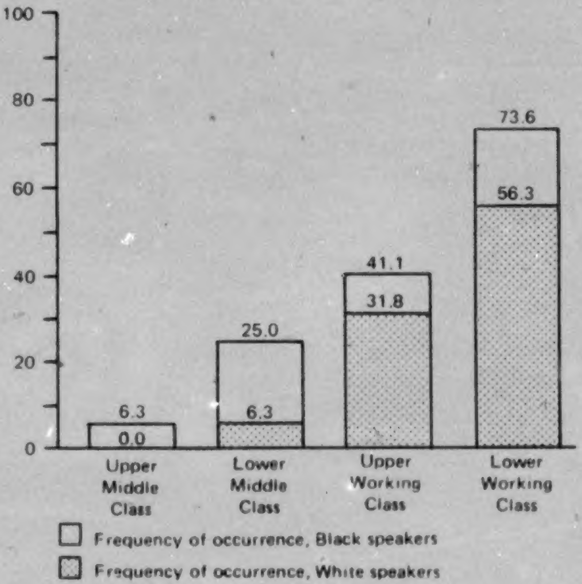
In short, then, the newer focus on dialect variability tended to build on the shoulders of previous linguistic work, adding the dimensions of a finer sampling procedure (random or stratified, rather than mere convenience sampling), an emphasis on grammar and phonology (as opposed to lexicon), a focus on variability and quantitative data (in contrast to single occurrence representation), and a sense of the primacy of the social group (rather than regional area) as the unit for correlation with linguistic variation.

Linguists were not satisfied however, with merely using more quantitative approaches to data gathering. William Labov (1969) claimed as a major goal in linguistics the need to incorporate such variability into the rules of grammar. This was a shocking notion to most linguists, in whose opinion and tradition linguistic variability, particularly socially conditioned linguistic variability, was not a part of the grammar at all.

One major goal of variable rule analysis was the attempt to incorporate such variability into the main body of linguistic theory. Labov spearheaded this approach, attempting to learn just exactly how variation works in language. But he also was interested in discovering the limits of grammatical competence. He was of the opinion that there is no end to the writing of grammars since the form that the grammar takes is a set of quantitative, variable relations. To give an example of the type of rule Labov proposes, consider the rule for contraction in Vernacular Black English which he constructs as follows:

FIGURE 1

Multiple Negation: Frequency of Occurrence in Detroit, by SES Group



Source: Shuy, Wolfram and Riley, 1967:88

$$\begin{bmatrix} +\text{voc} \\ -\text{str} \\ +\text{cen} \end{bmatrix} \longrightarrow \phi \;/\; \begin{bmatrix} {}^{*}\text{Pro} \\ \text{V} \end{bmatrix} \#\# \begin{bmatrix} \underline{\qquad} \\ -\text{T} \end{bmatrix} \begin{matrix} \underline{\text{Cl}} \\ \text{O} \\ \begin{bmatrix} {}^{*}\text{nas} \end{bmatrix} \end{matrix} \begin{bmatrix} \alpha & \text{Vb} \\ \beta & \text{gn} \\ -\gamma & \text{NP} \end{bmatrix}$$

This rule operates on a form in which the vowel has been reduced to a schwa; for example *he is* becomes he[αZ] with this rule making the form, *he's*. More technically, the rule deals with the removal of a schwa (+voc, -str, +cen $\longrightarrow \phi$) which occurs initially before a single consonant ($\frac{\text{Cl}}{\text{O}}$) in a word with a tense marker (−T) incorporated. When a pronoun precedes (Pro) or a nasal consonant follows (nas), the rule is categorical (*).

Variable rule analysis as constructed by Labov (1973) not only mentions the various alternative possibilities (structured grammar did as much, but swept some variations under the rug by calling them free variation), but also ranks how they constrain the rule. That is, when a form can optionally undergo a certain process, parts of the environment which influence the likelihood of the rule applying are identified and ranked. In this case, the alpha (α), or greatest constraint, does not show a high degree of ordering in that a preceding vowel (αV) and a following verb (αVb) have approximately equal effect on the application of the rule. The effect of a "gonna" (gn) following is less than either of these, however, and is therefore given the beta (β) constraint. The gamma (γ) constraint, the presense or the absence of a noun phrase, is even less powerful.

This ranking of the constraints on a variable rule grows directly out of quantitative analysis. Older, more traditional grammars often hinted at frequency statements or rule ordering constraints, but presented them vaguely if not sloppily. For example, note an older description of Walapai: "The dental and glottal fricatives are usually voiceless except that O is very often voiced intervocally and between a voiced consonant and a vowel" (Redden, 1966).

In contrast, quantitative analysis permits more precise observations. For example, English word-final stop consonants may be deleted, but the likelihood of deletion occurring is affected by both social constraints (age, sex, region, status, ethnicity, style) and linguistic environment (whether the following word begins with a vowel or a consonant and, if the latter, which type of consonant). Thus words like *and, bend, last,* and *first* are realized as *an', ben', las',* and *firs'* among blacks in Washington, D.C. Fasold (1972:67-70) found that the last consonant in *wild* and *east* when followed by words beginning with a vowel (*wild elephant* and *east end*) are deleted 28.7% of the time whereas the same consonant is deleted at the rate of 75.6% when followed by words beginning with a consonant *(wild horse, east precinct)*. Deeper investigation revealed that the deletion rule is even more favored if the first consonant in the cluster is a sonorant (a nasal or an *l*) and less favored when that consonant is a fricative or a stop. Thus the /d/ in sand castle is deleted 86% of the time, while the /t/ in *fast car* is deleted only 43% of the time.

Pure quantitative analysis, then, tells us that the consonant deletion rule is favored: (1) By not having a vowel immediately following; (2) By having a sonorant consonant rather than an obstruent (non-sonorant) consonant preceding. What is not clear is how to determine which of these two constraints on the deletion of the consonant outranks the other. A simple quantification of cases in which one factor favors the rule and another does not reveals the following (Wolfram and Fasold, 1974:103).

Environment	Example	% Deleted
following vowel, preceding obstruent	lif(t) it	25.2%
following vowel, preceding sonorant	wil(d) elephant	34.9%
following non vowel, preceding obstruent	fas(t) car	68.8%
following non-vowel, preceding sonorant	san(d) castle	83.3%

Thus the lowest percentage of deletion is found where neither feature favors deletion, and the highest percent where both features favor it. Equally interesting, in the middle cases in which the two features conflict, the higher percentage of deletion is where the non-vowel (favoring deletion) occurs. It appears, then, that the following non-vowel exceeds the effect of the preceding sonorant. Therefore, a following non-vowel is a stronger constraint than a preceding sonorant.

Educational Implication of Quantitative Analysis

What significance might this have to educators? Several potential applications seem to emerge.

First, *structural integrity*. Quantitative analysis displays, more than ever before, that *language is rule governed*. While the extent of knowledge teachers need about this principle is uncertain, it has always seemed to me that we should have answers available for teachers and children when they ask.

Second, *diagnosis*. Quantitative analysis can pinpoint the exact focus for teaching. A teacher who knows that vernacular speakers do something strange with the ends of words is in only slightly better position than one who knows nothing at all about vernacular. A teacher who realizes that what happens relates to consonant clusters and that it is the second consonant which deletes is in a better position to diagnose and prescribe. A teacher who realizes that in order for consonant cluster reduction to take place, both consonants must share voicing is in even better shape since this prevents the teacher from worrying about the wrong items, such as *belt* or *wart*. A teacher who knows something about the effect of linguistic environment (following vowel, preceding sonorant, for example) will be able to anticipate the diagnosis even better. Whether or not all teachers need to know all of this is debatable, but quantitative analysis has begun to make it possible.

Third, *prediction*. Quantitative analysis can predict learning sequences and permit the teacher to determine where a learner is in the acquisition of language. Table 1, from Wolfram (1969), relates to the acquisition of standard English, but similar examples might be cited in early child language acquisition as well. Not surprisingly, the highest output frequency of the rules is in the lowest classes and all four classes have the three rules ranked in the same order based on frequency of occurrence. Thus, the pattern is consistent across class even though the frequencies vary, suggesting that teaching to these rules is useful to all groups but to a lesser extent to the upper SES groups.

TABLE 1

Interaction of SES with Three Stigmatized Rules
In Detroit Vernacular Black English

Class	Rule: θ f(bath baf)	Rule: r ∅ (car cah)	Rule: stop ∅ (band ban)
Upper Middle	.06	.21	.51
Lower Middle	.11	.39	.66
Upper Working	.38	.61	.79
Lower Working	.45	.71	.84

Probability and Variation

Labov's variable rules are written for specially well-defined and previously determined social groups and are based on the frequency of occurrence of the feature under specific conditions. Henrietta Cedergren and David Sankoff (1974) adopt basically the same approach but bring a more sophisticated mathematical theory to the task by using probabilities associated with rules rather than frequencies. They feel that a person's performance is a statistical reflection of his competence. The frequencies observed in individual performance are used to determine the probabilities that each constraint, whether linguistic or social, contributes to the application or nonapplication of a particular rule. Naturally, such precise numbers do not exist in the heads of speakers; rather, statistical tendencies are what is reflected. In such a manner, rules are written for the speech community, and these rules specify the linguistic constraints on their applications. They are accompanied by tables which provide the probabilities for each of the linguistic constraints and for any relevant social parameters.

In an effort to test the appropriateness of this approach, Cedergren and Sankoff performed an experiment on r-spirantization. Using the probabilities determined by the speech community and for an individual's social class (which turned out to be the significant social constraint), the researchers checked the match between the predictions made by the rule and the observed data for each individual. The predictions turned out to be fairly close, confirming the hypothesis that the rule for speech community accounted for the performance of individual members. This equal use of social parameters and linguistic constraints to account for language variation, then, operates somewhere between the extremes of social constraints as primary and linguistic constraints as the independent variable.

Implicational Scaling

In order to discuss the primacy of the linguistic constraint in the study of language variation, it is first necessary to describe a linguistic method known as implicational analysis. Although implicational scales have been used in other disciplines (especially in sociology, where they bear the name of Guttman scales), they are relatively new to linguistic analysis. David DeCamp (1971) began to experiment with such scales as he worked with Jamaican creole, and the approach also has been used by linguists on various social dialects in the Americas (especially Bickerton, 1972, and Wolfram, 1974).

C.-J. N. Bailey (1973) is a prominent advocate of the "linguistic-constraint-as independent-variable" philosophy of language of variation. His goal is to write panlectal rules which cover the entire language system. Each individual has a subset of the rules and more general forms of the rules than the panlectal rules which account for them. A speech community, in this case, is a group of people who evaluate linguistic variables in the same way (as favored or as stigmatized) and who have the same patterns for the usage of these variables.

Implicational scales are used in rule writing in such a way that a pattern of outputs is implied in the rule itself. Bailey maintains that the time factor accounts for all other kinds of differentiation, whether geolographical, social stylistic, or whatever. Thus his rules include the notions of marking (based on further developments of the phonological marking of Jacobson [1968] and Chomsky and Halle [1971]) and implicational coefficients in such a way that the rule generates an implicational pattern of outputs which also take into consideration the environments in which the outputs occur. This series of outputs makes up a series of temporally differentiated lects which are minimally different from those which follow (called isolects). This temporal differentiation reflects the social parameters of language, according to Bailey, who goes on to treat them as algorithms which define the place in the series of temporal isolects where a particular combination of social characteristics falls.

Thus these algorithms are devices which convert unilinear implicational patterns into multidimensional sociolinguistic matrices. The relevant social parameters are probably best identified by trial and error, as Fasold, Wolfram, Labov and others have done with the variables of social class, race, sex, style and age. In considering the dynamic aspects of language, age factors seem to be the most obvious differentiations, but this need not always be true. If a given rule has four environments, in such a way that environment 1 is heavier-weighted than environment 2, which is more than 3, which is more than 4, the implicational output will generate the application of the rule first in 1 and last in 4. Since 4 is the lightest-weighted environment, its presence implies the presence of all heavier environments.

In Vernacular Black English, for example, the rule for *t, d* deletion in a particular linguistic environment may be described in a multi-dimensional sociolinguistic matrix at one particular time as Figure 2 demonstrates.

FIGURE 2

Implicational Scale Indicating Language Change in Progress

	Most Informal	Rather Informal	Rather Formal	Most Formal
Upper Middle Class	2	1	--	--
Lower Middle Class	3	2	1	--
Upper Working Class	4	3	2	1
Lower Working Class	†	4	3	2

† = Categorical application of the rule
-- = Nonapplication of this rule

The change here is seen to have begun in the lowest class in informal speech. The wave-like characteristic of the outputs is clearly indicated. Sociolinguistic algorithms can be used to determine what temporal isolect is used by a person with certain social characteristics when the isolect associated with one set of characteristics is known. For change involving disfavoring, an algorithm might state that one isolect is less advanced for each more monitored style. In this way, the linguistic aspects are treated as central, and a rule can be written to generate temporal differentiation which will then fit the social differentiation (keeping in mind that in this model, various types of social differentiation are embraced as temporal differentiation in language change).

Wolfram and Christian (1975) have noted several quantitative problems inherent in the use of implicational scaling. One difficulty is in aggregating the informants into different varieties since there seems to be no principle for deciding whether a person fits into †,4,3,2,1, or --. The pattern, of course, in Figure 2 is ideal.

As this discussion highlights, linguistic analysis has begun to be affected by quantitative approaches. Out of an initial concern for social dialects has developed a mission to the field of linguistics itself, a mission which has opened the doors of inquiry considerably wider than when the only legitimate concerns of linguistics were for abstract universals. This newer focus has clearly demonstrated that the concern for variability is not mere surface level triviality and that human society must be considered along with the human mind as we examine the fantastic complexity we call language.

Generalizing and Grouping Speakers and Writers

In contrast with the legitimate use of quantitative data for analyzing the language used by people is the less legitimate use to quantify the people who use language. The former displays facts which might otherwise be obscured; the latter groups people who might otherwise be considered different and whose differences might be critical for accurate analysis.

What this means is that quantification can be used in two exactly opposite ways: (1) micro-analysis—to reveal information which is useful for understanding an individual's use of language, direction of language change, or stage of learning language; and (2) macro-analysis—to obscure an individual's variable use of language in order to fit him into a general category with other people whose variability is similar.

Linguists can be guilty of using quantitative data to obscure differences, as even Figures 1 and 2, and Table 1 can be taken to illustrate. Language normally operates on a continuum rather than on four-point scales; when we segment a number of people into classes, we are actually obscuring differences which, in individual cases, might be diagnostically important. Thus quantitative analysis in linguistic studies, as in any other field, can be used both to probe for deeper patterns and to gloss over differences for more general or homogeneous groupings. If this seems paradoxical, it should not be surprising, for a great deal of language behavior is paradoxical.

Speakers, for example, must be enough like each other to be understood while, at the same time, being different enough to establish their own personal identity and give clues about their group memberships (age, race, sex, SES, etc.). These differentiating functions of language are little understood by linguists, much less by educators, and they are basically unresearched. The need for a third grade boy to establish his sense of masculinity for example, has been known to affect his willingness to "read with expression" in his reading group (Shuy, 1977). We know very little about how this differentiation process works or the extent to which it is consciously done. In fact, we have been so concerned with similarity-finding in language analysis that we have neglected such obvious and interesting topics as the effects on language of institutions or occupations (what is it like to talk like a lawyer, an airline stewardess, or a teacher?). We certainly can benefit from examining the speech in the communication exchange between doctor and patient; indeed, recent research has revealed that the major problem in such communication resides in the physician, not just the patient (Shuy, 1975). In all of these areas, fruitful research of both a quantitative and qualitative nature can be expected in the near future.

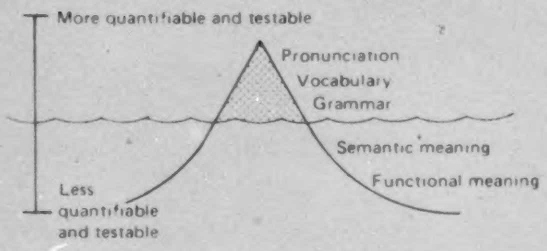
Some Problems in Language Measurement

The more linguists study the semantic and pragmatic meaning conveyed by language, the less comfortable they become about the possibility of accurate measurement by tests which use language as a medium. It is beginning to be believed, in fact, that the most critical measurement points of all, at least as far as language is concerned, are the ones least susceptible to quantification.

In language teaching, for example, it is considerably easier to measure accuracy in pronunciation or vocabulary than in meaning. As shown in Figure 3 the typical evaluation points in language measurement may be plotted like an iceburg with the visible features above the water line but the more critical ones below (Shuy, 1976).

FIGURE 3

Visible Evaluation Points of Language

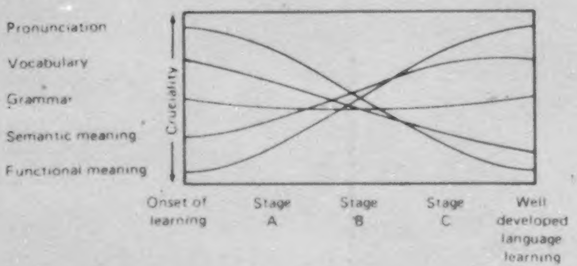


If one were to construct a new teaching program for learning a language, one could probably be persuaded that the most important activities involve getting meanings across to another person. In doing this, one also adds or detracts from personal effectiveness by using or failing to use appropriate vocabulary, pronunciation, and grammar. Likewise, one could probably make a good case for hierarchies of importance even *within* the categories. It might be reasonable, for example, to assume that it is more important to be accurate in one's past tense markers than in subject-verb agreement. A really good language learning program would probably note the occurrence of all of the variables in Figure 3, perhaps even in some sort of dynamic framework as in Figure 4 (Shuy, 1976).

Figure 4 suggests that at the early stages of language learning, pronunciation, vocabulary, and grammar are more critical measurement points of learning ability than they are at later stages and that the categories of semantic and functional meaning, though present at the onset of learning, become increasingly important as a learner progresses toward well-developed abilities. This same schematic drawing may also illustrate the dynamic of learning to process language in reading (Shuy, 1975), as shown in Figure 5.

FIGURE 4

Tentative Language Program Dynamic



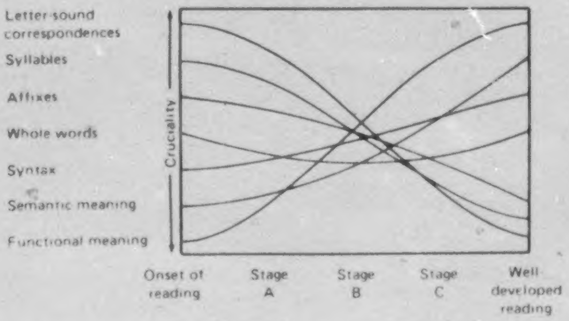
Typically what happens is that the more visible, highly recurring features (vocabulary, pronunciation, and sometimes grammar for language learning and letter-sound correspondences and whole words or decoding strategies for reading) are measured and quantified throughout the learning program *without regard for their interrelationships with the other accompanying measurement points.* These features are measured because we know what they are and because their inventory is reasonable to assess more than because of what they tell us about language learning or reading ability. Because they recur, they are easily quantifiable, thus lending an air of scientific respectability. When couched as test questions, such features become discrete point test items, and it is assumed that by knowing the answers to such questions, one evidences significant ability in the gestalts of language or reading.

Once one has progressed dubiously this far, the next easy step is to assume that because a feature measures something meaningful at one stage, it can be measured continuously for that same meaning. The absurdity of such an assumption is particularly apparent in reading where the usefulness of measuring letter-sound correspondences as an indication of reading ability becomes more and more doubtful as the learner becomes less and less conscious of this skill. Some early learning skills are important at some stages but potentially harmful to keep in conscious awareness beyond those stages.

Any attempt to quantify a child's ability to respond to questions related to letter-sound correspondences after that child has progressed to rather advanced stages of reading ability, therefore, runs the risk of

FIGURE 5

Possible Reading Program Dynamic



measuring an irrelevant ability—one which is necessarily being submerged in the child's consciousness in favor or more cognitive later-learning strategies. If advanced children continue to do well on such efforts to measure their ability, the gains may be only a result of their ability to become good tests takers as well.

Measuring Functional Language: A Case Study

This section describes some current research in measuring functional language. In doing so, quantification is far from the core of our thinking since scientific measurement in this area is not dependent on it. Experiences growing out of the Lau vs. Nichols Supreme Court decision, the Aspira Consent Decree in New York City, and the various bilingual education bills have revealed a basic gap in the knowledge base for educating children whose dominant language is not English. There is no doubt that legislative and judicial action has effectively provided momentum to make education more responsive to the needs of these children, but the momentum *requires educational technology that is only beginning to be developed.*

For example, the Aspira Consent Decree requires that the placement of children in educational programs using English or Spanish as the medium of instruction be determined by their ability to *effectively participate* in the instruction. This legislation precedes by a wide mark the technology upon which it can be based. No assessment instruments are available which purport to test this ability. There is a growing concensus among second language specialists that tests of grammar and phonology are not accurate predictors of effective participation and that functional language competence is far more crucial. That is, a child's ability to seek clarification or get a turn seems much more crucial than his ability to use past tense markers properly. To develop the necessary assessment instruments requires an inventory of the functional language competence demanded in the educational setting at the various age/grade levels.

Functional language competence is the underlying knowledge that allows people to use their language to make utterances in order to accomplish goals and to understand the utterances of others in terms of their goals. It includes a knowledge of what kinds of goals language can accomplish (the functions of language) and of what are permissible utterances to accomplish each function (language strategies). Table 2 displays a small sample of some functions, strategies, and utterances for adult English speakers.

Table 2 is in no way complete. There are many more functions, many other strategies for each function and, of course, many other utterances which could be used for each strategy. More important, the table is incomplete in that the context of each utternace needs to be specified to insure that the utterance is permissible to accomplish the function.

TABLE 2

Sample of Functional Language Knowledge
for Adult Speakers of English

Function	Strategy	Utterance
Giving an order	Performative	I hereby order you to come home.
	Direct Imperative	Give Jane some food.
	Wh-Imperative	Won't you please buy me some candy?
	Statement	Mr. Jones, I need some more paper.
Promising	Performative	I hereby promise you that I will be home by eleven.
	Future Statement	I'll be home by eleven.
	Conditional Statement	If you give me a dollar, I'll be home by eleven.
	Question	Will you let me take care of my own affairs?

Functional language competence also accounts for knowing what utterances cannot do. In English, the statement, "*You are fired,*" works to fire the addressee but the utterance, "*You are a frog,*" does not work to turn the addressee into a frog. Likewise if a teacher tells a student, "*You have one minute to get over here,*" the utterance can act as an order, but if the student says the same thing to the teacher such a meaning is, at best, farfetched.

This very sketchy discussion of some aspects of functional language competence shows that a speaker's underlying knowledge must be extensive and complex. In the literature of *linguistics, sociolinguistics, and philosophy,* three other terms are also used to refer to functional language competence: *communicative competence/pragmatics of natural language/speech act competence.* All who have studied this phenomenon agree that language users cannot possibly learn and store in memory all the complexities of functions, strategies, and utterances as item lists any more than they can store phonological or grammatical language as item lists. This knowledge must be learned and stored according to organizational principles. These principles may be considered constitutive rules which account for the successes and failures in the utterances meant as promises, for example, but they also separate promises from orders, requests for information, etc. In a similar manner, the constitutive rules of football not only account for the successes or failures of particular plays but also account for football and not baseball or basketball.

It appears that language functions, unlike phonology and grammar, are developmental almost throughout life. Few adults, for example, ever become proficient at the language function of condoling. For the sake of survival, children learn rather early how to interrupt appropriately. One also learns how to avoid being interrupted, how to get or avoid a turn in talking, how to refuse, how to clarify, how to obfuscate with dignity (see especially Watergate transcripts). What may be considered rudeness may only be an imperfectly developed sense of interruption skills. It would seem critical that teachers be able to distinguish these matters.

Questions About Quantification in Measuring Functional Language

Theoretical discussions of conversational rituals and routines, politeness, the organization of discourse, implications invoked by language, presuppositions, illocutionary acts, and perlocutionary acts serve as a partial background for the study of functional language. This interplay of theory and data, each informing the other, is another hallmark of linguistic descriptions. Part of this interplay can be seen in three aspects of CAL/Carnegie's recent work on functional language development which bear on questions about quantification. These questions are concerned with what to count, overlapping relationships, and generalizing.

The first question is concerned with determining what to count and what not to count as an instance of a category. Labov identified this as the major part of a linguistic analysis. For example, there are certain ways to address people which indicate that the speaker thinks of them as of higher status. This is reflected in some languages in the pronominal system where there are multiple forms for "you" and in current American English by varying use of titles and first and last names. However, there are occasions when one uses address terms not because the speaker-hearer relationship has certain status definitions, but rather because the status definition is being made by the speaker or being pretended or even mocked. So the following utterance by a teacher of four-year-olds to a four-year-old dressed in a suit and tie cannot be counted the same way as the more regular address term utterances: "Well, Mr. Bobby Johnson, you certainly look handsome today."

These special uses might in fact be considered "metalinguistic" in that to have an effect they depend on the existence of the regular rule for using "Mr." and on that rule excluding these very cases. Such special cases have been cited concerning functional language (see especially Erickson and Shultz, 1976). It may be that these uses play a role in the category they are formally or structurally associated with as well as being members of a special metalinguistic category that needs to be added. A careful analysis that recognizes the interplay between data and theory is called for. Without it, simple quantification would be misleading or useless.

The second question concerning quantification grows out of the familiar fact that paradigmatic and syntagmatic language relationships often overlap. Paradigmatic relationships hold among language elements that occur in similar places in utterances. One item is used and the others are not (the personal pronouns are paradigmatically related, for example, and "he" is used rather than "she" under certain conditions). Syntagmatic relationships hold among language elements that occur together in utterances. (The agreement of a singular subject and a present tense verb is an example.) In functional language, we might see the following utterances paradigmatically related as ways to give a command, to get the addressee to perform an action:

1. Raise your hand, Mark.
2. Would you raise your hand, Sophia?
3. In this class we raise our hands, Gene.
4. I can't hear you because you didn't raise your hand, Gene.

However, if we look at stretches of discourse for the syntagmatic relations, we can see that similar utterances may also be syntagmatically related. A teacher in a first grade class can use the following utterances in sequence.

1. I can't let you line up for playing because the magic markers aren't back where they belong.
2. We have to make sure that the magic markers are back on the shelf.
3. Can you put the magic markers back?
4. Put the magic markers in the box on the shelf.

(This display should not be taken to mean that these utterances do not also have syntagmatic relationships with the utterances and actions that occur between them, even if uttered by someone other than the teacher.)

The range of facts available as data in categories like those listed above must all be accounted for by a description and explanation of functional language while, at the same time, the description and explanation may display that more facts need to be considered. The particularly interesting aspect of the interplay between theory and data in functional language is that the same elements can be in both a syntagmatic and paradigmatic relationship to each other as shown above. In such cases, simple quantification must be interpreted carefully.

Still another question concerning quantification is characteristic of all naturalistic data collection. No airtight argument is possible for generalizing from a variety of videotaped episodes to the language competence of a child or of children in general. This is the "How do you know it didn't happen when the camera was off?" controversy. We do not hope to overcome this with a statistical probe but rather by using our naturalistic data to form a falsifiable hypothesis which can then be tested on our population, i.e., we extend our corpus to insure inclusion of the crucial cases postulated by hypotheses.

Corpus extension techniques have been used by a variety of linguists. Berko-Gleason devised a technique to determine the child's level of acquisition of morphological principles. Fasold (1972) devised techniques to investigate underlying syntactic forms. Carden (1973) and Elliot, Legum, and Thompson (1969) elicited more directly from adults. Quirk and Svartvic (1966) devised a method to blind subjects to the real tasks to gather data on a variety of structures. What all of these have in common with our effort is that the investigator structured an environment so that subjects would perform tasks postulated to reveal information about the subjects' tacit knowledge of his language. Our effort differs as follows.

First, like Berko-Gleason, we are directly investigating production ability but we are simultaneously testing and formulating hypotheses specific to our task. Second, like Fasold, we are investigating underlying aspects of language but we are not relying on the technique described here to establish discrete underlying units. Third, like Carden and Elliot, Legum, and Thompson, we are investigating directly but not at the sentence level. Fourth, like Quirk and Svartvik, we are "blinding" our subjects to the real task, i.e., to demonstrate their ability to use language structures to accomplish social tasks, but we are using social situations rather than grammatical operations as the distractor.

The problems in tests and interview situations—including mismatches in language structures, background assumptions, and task identification between the test constructor and the test taker—have been pointed out by a variety of linguists, sociologists, anthropologists, and reading specialists (cf. Shuy, 1976; Wolfram and Griffin, 1974; Cicourel, 1974).

A basic problem is that the goal of getting responses that will be comparable across subjects or across testing times is often realized by forcing one standard interpretation of a question (or stimulus) and answer (or response) that is, in fact, not uniquely interpretable but rather is vague and can be fully specified only with reference to specifics of the individual test-taker's background and the individual test-taking occasion.

Test problems recognized, we still need to have a way of controlling some of the variables in functional language utterances and of eliciting special forms that may have eluded naturalistic observation techniques. It is not our intention to detail here the corpus extension technique used in the CAL study. Suffice it to say that it was carefully constructed to get language information from children in a way which seems to allow for analysis across children without requiring that the children be the same as each other or the test-taker and which provides results that allow for the interplay between theory and data needed for functional language analysis. Whether this kind of technique can be used in large-scale assessment of children or for diagnostic purposes is yet to be determined. We feel it avoids to a large degree the problems we have noticed in tests, but it may be that psychologists and test constructors can identify some new and insurmountable problems in it. At any rate, for our current purposes it gives us data that we can quantitatively and qualitatively analyze to deal with the basic issues raised above.

Conclusions

This paper has intended to assert that quantitative analysis was a welcome additon to linguistics, especially at a time when attention turned to variability in language. Quantitative studies have enabled linguists to probe more deeply into the structure of language, particularly with regard to frequency of occurrence of certain features (and the effects of such work on rule ordering constraints) and into the issue of probability in language production. The implications of such study for education are largely in the areas of individual diagnosis and placement.

Quantitative analysis is less comfortable to linguists when it is used to generalize or obscure linguistic differences. The latter seems to be the more common use of quantitative data in education. Linguists worry a great deal about various difficulties posed by quantification, especially when such measurements treat the less significant elements of language or fail to take note of the dynamic nature of language.

Linguists have not yet solved the question of how much to sample. It appears that language data are relatively undefiled by conscious awareness and therefore susceptible to smaller samples than purchasing patterns, voter preferences, etc. Linguists also are concerned about the meaningfulness of non-occurrences of linguistic features. Since the inventory of possibilities is so great, it is necessary to know the relevance of the lack of occurences of the element and to note the linguistic and social location of its occurrences. Nor has the last word been said about how to establish thresholds for language stratification, dialect boundaries, etc.

A great concern exists among linguists that care be taken to quantify like elements and that these elements be identified for what they are. One of the greatest accusations made by linguists against standardized language tests is that they do not measure what they say they measure—that they do not have content validity. Naturally, we also worry about the meaning of empty cells or non-patterning in our analysis of language, particularly if we suspect that such surprises are artifacts of our coding schema or analytical mode.

Perhaps the conclusion of a paper such as this is the appropriate place for a plea for caution in the use of quantitative data in language when the analysis moves outside the range of the study of an individual's speech. Quantification is much safer when limited to diagnostic concerns. What is more, a huge world of exciting research exists at such a micro-level. It seems to be time to stop thinking of large N's and to start analyzing the language abilities of a few people with greater intensiveness and depth.

Notes

1. For the most useful history of linguistic theory, see: D. Hymes. *Studies in the History of Linguistics*. Bloomington: Indiana University Press, 1974.
2. These figures represent a number of informants in each of the four SES groups and a large quantity of occurences of the feature for each informant represented in the group. In the case of multiple negation, in addition to tabulating the occurrences, it was necessary to see them in relationship to a meaningful touchstone. Thus, every single negative and every multiple negative in each speaker's speech sample were added together to form a universe of potential multiple negatives. These figures, then, display the relationship of the occurrence of multiple negatives in relationship to all potential multiple negatives.

References

Bailey, C.-J.N. *Variation and Linguistic Theory*. Arlington (Va.): Center for Applied Linguistics, 1973.

Bickerton, D. "The Structure of Polylectal Grammars." In R.W. Shuy (ed.). *Sociolinguistics: Current Trends and Prospects*. Washington: Georgetown University Press, 1972.

Carden, G. "Disambiguation, Favored Readings, and Variable Rules." In C.-J.N. Bailey and R.W. Shuy (eds.). *New Ways of Analyzing Variation in English*. Washington: Georgetown University Press, 1973.

Cedergren, H.J., and D. Sankoff. "Variable Rules: Performance as a Statistical Reflection of Competence." *Language* 50:333-55, 1974.

Chomsky, N. and M. Halle. *Sound Patterns of English*. New York: Harper and Row, 1971.

Cicourel, A. "Some Basic Theoretical Issues in the Assessment of the Child's Performance in Testing and Classroom Settings." In A.V. Cicourel, et al. (eds.). *Language Use and School Performance*. New York: Academic Press, 1974.

DeCamp, D. "Toward a Generative Analysis of a Post-Creole Speech Continuum." In D. Hymes (ed.). *Pidgnization and Creolization of Languages*. Cambridge: At the University Press, 1971.

Elliott, D., S. Legum, and S. Thompson. "Syntactic Variation as Linguistic Data." *Chicago Linguistic Society* V:52-59, 1969.

Erickson, F., and J. Shultz. "Social Solidarity and Indexical Repair." (Ms.), 1976.

Fasold, R. *Tense Marking in Black English*. Washington: Center for Applied Linguistics, 1972.

Jacobson, R. *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton, 1968.

Labov, W. *The Social Stratification of English in New York City*. Washington: Center for Applied Linguistics, 1966.

__________ "Contraction, Deletion, and Inherent Variability of the English Copula." *Language* 45:715-62, 1969.

__________ *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1973.

Quirk, R., and J. Svartvik. *Investigating Linguistic Acceptability*. The Hague: Mouton, 1966.

Redden, J. "Walapai I: Phonology." *International Journal of Applied Linguistics* 32:1-16, 1966.

Shuy, R.W. "The Patient's Right to Clear Communication in Health and Mental Health Delivery Service." Paper presented at the Conference on Health and Mental Health Systems, Philadelphia, November, 1975.

__________ "Pragmatics: Still Another Contribution of Linguistics to Reading." In S. Smiley and J. Towner (eds.). *Language and Reading*. Bellingham: Western Washington State College, 1975.

__________ "Implications of Recent Sociolinguistic Research for the Problems of Migrant Worker Children." Ghent (Belgium): A.I.M.A.V., 1976.

__________ "The Mismatch of Child Language and School Language: Implications for Beginning Reading Instruction." Paper presented at Conference on Beginning Reading Instruction, LRDC, Pittsburgh, April, 1976.

__________ "Sociolinguistics." In R. Shuy (ed.). *Linguistic Theory: What Can it Say About Reading?* International Reading Association, 1977.

Shuy, R. W., Wolfram, W., and Riley, W.K. *Field Techniques in an Urban Language Study*. Washington: Center for Applied Linguistics, 1968.

Williamson, J.V. "A Look at Black English." *The Crisis* 78:169-173, 185, 1971.

Wolfram, W. *A Sociolinguistic Description of Detroit Negro Speech*. Washington: Center for Applied Linguistics, 1969.

__________ *Sociolinguistic Aspects of Assimilation: Puerto Rican English in New York City*. Washington: Center for Applied Linguistics, 1973.

__________ "The Relationship of White Southern Speech to Vernacular Black English in the Deep South." *Language* 50:498-527, 1974.

Wolfram, W., and D. Christian. *Sociolinguistic Variables in Appalachian Dialects*. (Final report to NIE) Arlington (VA.): Center for Applied Linguistics, 1975.

Wolfram, W., and R.W. Fasold. *The Study of Social Dialects in American English*. Englewood Cliffs (N.J.): Prentice-Hall, 1974.

Wolfram, W., and P. Griffin. "A Sociolinguistic Analysis of the Armed Services Vocational Aptitude Battery." (Final Report. D.O.D. Contract No. DAHC1573CO364), 1974.

LANGUAGE ASSESSMENT: WHERE, WHAT AND HOW

Courtney B. Cazden
Harvard University

James T. Bond, Ann S. Epstein, and Robert D. Matz
High/Scope Educational Research Foundation

Sandra J. Savignon
University of Illinois,
Champaign-Urbana

For better or worse, there is a growing trend to include assessment of oral language in educational evaluation. In California, schools receiving state money for an early childhood education program must include and evaluate an oral language component. Several Follow Through sponsors, concerned that traditional measures of school achievement fail to tap their program strengths, are collaborating to develop both oral and written language measures. Bilingual education programs will be pressed to document their effectiveness in increasing competence in the second language at least, and hopefully in the first language as well.

This could be a welcome trend. Evaluation of any aspect of a school program becomes a symbolic validation of the expenditure of school money and teacher time on that aspect. To put the matter bluntly, the evaluation of oral language could itself exert pressure against silent classrooms. Or if the evaluation is ill conceived, it could be a disaster. Sometimes funding decisions may really be made on the basis of evaluation data; even when that does not happen, evaluation instruments become an implicit in-service curriculum for teachers, an internalized analytical framework that influences the mini-tests teachers continuously construct in the classroom as they take children's words as indicators of what they have learned (see Cazden, 1976a).

What kind of analytical framework do we want teachers to internalize? On what kinds of evidence about oral language do we want funding decisions based? What is the proper role of numbers in accumulating that evidence? Although any complete evaluation would assess both receptive and productive language, only the latter will be discussed here. Productive language is where both recent innovative efforts and the most severe problems are.

Two questions to be answered concern the "where" and the "what": decisions about the assessment situations and about focal aspects of communicative competence to be assessed in them. Neither is a new question. In the discussion here, two examples of current work will be inserted, from papers prepared for this purpose by the staff of the High/Scope Educational Research Foundation and by Sandra Savignon of the University of Illinois. Following these two sections, "how" questions about the roles of quantitative and qualitative evidence are included in a more general exploration of the differences between program assessment and the diagnosis of both children and classroom environments.

The Assessment Situation

One critical problem for evaluation design is sharply posed by Shapiro's experience with "a pilot effort to try out promising techniques for studying young children who had and had not participated in a Bank Street-sponsored Follow Through (FT) program" (1973:256). According to Shapiro, "When we observed the children in their classrooms [low-income Black children in six first grades], there were striking differences between the FT and comparison classes; when we compared the children's responses in the test situation, there were no differences of any consequence" (p. 527).

This is not an uncommon result. But its importance in this case is enhanced by the care with which Shapiro designed the tests. She used six different techniques "to provide a range of measures and to offer the children some variety in task requirements": general interview questions, sentence completion items, Draw-A-Person, a self-rating technique, and two techniques adapted from Wallach and Kogan—Instances of a Category and Line Drawings. She analyzed the responses in both qualitative and quantitative terms.

Retrospectively, when the analyses showed no significant differences attributable to the program, Shapiro reconsidered her evaluation design (1973: 533).

> What children do in the classroom—the kinds of questions they ask, the kinds of activities they engage in, the kinds of stories, drawings, poems, structures they produce, the kinds of relationships they develop with other children and the teacher—indicates not only what they are capable of doing but what they are allowed to do. Classroom data are generally downgraded in attempts to study the effects of educational programs because we cannot know whether the comparison group, given the same opportunities, would behave in similar ways. And conversely, we do not know whether, if the opportunity were removed, there would be any carry-over to a new classroom situation, that is,

> whether the effects have been internalized. Nor is it easy to separate the contribution of and effect upon individual children in the group. Following the line of reasoning of an earlier study, I assumed that the internalized effects of different kinds of school experience could be observed and inferred only from responses in test situations, and that the observation of teaching and learning in the classroom should be considered auxiliary information, useful chiefly to document the differences in the children's group learning experiences.
>
> The rationale of the test, on the contrary, is that each child is removed from the classroom and treated equivalently, and differences in response are presumed to indicate differences in what has been taken in, made one's own, that survives the shift to a different situation.
>
> The findings of this study, with the marked disparity between classroom responses and test responses, have led me to reevaluate this rationale....

Shapiro's dilemma is deeply troublesome. The kind of evidence accepted as proof of good education seems to change sharply depending on whose children we are making decisions for—"ours" or "theirs." When we as parents select a school for our own children, we go as observers to see how children and teachers spend their time and probably give less weight to evidence from test scores. But when we as researchers or government officials have to make decisions about "their" children—nonwhite, immigrant or poor—the grounds of accountability shift so that only numbers in the form of test scores count. And then, in that numbers race, certain kinds of education almost always "win" and other kinds almost always "lose." The Pueblo classroom described by Vera John (in Leacock's paper) would probably lose, as did the classrooms for black children described by Shapiro.

Recommendations for alternative evaluation plans divide in two: design more valid test situations or rely more on evidence from ongoing classroom life. As we shall see, in actual practice, the two alternatives tend to merge.

Test Situations

A problem with most assessment situations, especially but not only for assessing oral language, is what Bronfenbrenner calls "ecological validity." Bronfenbrenner (1976:1-2) has argued at length that "much of contemporary developmental psychology is the sciece of the strange behavior of children in strange situations with strange adults for the briefest possible periods of time," and he extended this argument to educational research in an invited address to AERA in San Francisco. In his words, the first defining property in ecologically valid research is as follows:

> Proposition 1. An experiment is ecologically valid when it is conducted in settings that occur in the culture or subculture for other than research purposes, or might occur if social policies or practices were altered. Accordingly, in contrast to conventional experiments, in which the setting, participants, and activities are often unfamiliar, and the experiment is a one-time event of short duration, ecological experiments involve places, social roles, and activities that are enduring and known to the participants because they occur in everyday life. The requirements of *ecological validity* applies to all the elements of the setting, including the elements designated as the experimental outcome (1976:12).

Others have discussed similar contrasts. I have given the labels "concentrated encounters" and "contrived encounters" to more and less ecologically valid assessment situations (e.g., Cazden, 1975). Concentrated encounters are concentrated forms of real-life interactions, while contrived encounters are the traditional test situations, interactionally and motivationally impoverished as they usually are, in which we try to elicit oral language on demand. Especially for young children, assessment situations must be concentrated forms of interaction experiences familiar from the classroom. For older students, they can be concentrated forms of interaction situations that will be encountered on the job or as a citizen. The latter is what Gagne (1975:154) calls "job sample testing" that achieves relevance and validity "by representing the *stimulus situation* which matches that of the learning objective."

For examples of ecologically valid oral language assessments, I asked colleagues to contribute sections of this paper. Bond and his associates discuss their attempts over the years to develop language assessment procedures appropriate for their own educational programs for young children. Savignon describes how she assesses the communicative competence of college students learning a foreign language.

METHODS FOR ASSESSING THE LANGUAGE PRODUCTION OF THE YOUNG CHILD

J.T. Bond, A.S. Epstein and R.D. Matz

The High/Scope Educational Research Foundation not only conducts educational research but also develops and implements educational programs. The language assessment procedures described here have been developed primarily for program evaluation rather than for the formative-diagnostic evaluation of individual children or for basic research on language development. Moreover, they are designed to assess educational objectives derived from a particular educational philosophy.

High/Scope's approach to education grows out of cognitive developmental theory, particularly the work of Jean Piaget. Learning is a cooperative, social enterprise motivated by the interests and satisfactions of the learner rather than by teacher-controlled rewards. The process of education is active and generative rather than passive and responsive.

Traditional measures of language compentency have emphasized the *formal-mechanical* aspects of language performance, *correctness of usage*, and *decoding* rather than encoding. High/Scope's educational program and others stressing the active-generative role of the learner require different evaluation procedures. They require situations in which the subjects work with real objects to solve challenging problems of interest to them; situations that foster individualized, divergent behavior and encourage purposeful communication through language and other media. The measures described here are the interim products of long-term effort to develop evaluation procedures that meet these requirements in the assessment of children's productive language. Language analysis variables emphasize the *functional* aspects of language used in purposeful communication, and they are concerned with *meaning* and *effectiveness* rather than form and convention. Throughout, instrument development has been guided more by developmental theory and social philosophy than by psychometric expediency.

Productive Language Assessment Tasks (PLAT). The Productive Language Assessment Tasks (PLAT) measure children's ability to express their thoughts and feelings through written language (see Bond, 1975). The tasks allow children to work with real objects, structuring and solving problems of their own design. Social interaction is encouraged throughout. The tasks elicit written representations that are founded in immediate, concrete experiences and structured largely by the child rather than by stimuli associated with the measurement procedure. Although the PLAT is unabashedly curriculum-specific, the instrument taps dimensions of language behavior fundamental to general language competency.

The PLAT includes two tasks–reporting and narrating. In the reporting task, children are given raw materials and tools and asked to make anything they want. After about thirty minutes, they are asked to write about "how" they made whatever they made. Thirty-five minutes is allowed for writing reports, and children may interact with one another during both phases of the task.

In the narrating task, children are given sets of relatively unstructured materials to "help you make up a story." They are explicity encouraged to interact with one another and typically engage in intensive dramatic play. After about fifteen minutes, they are given about thirty-five minutes to write a "make-believe or pretend story."

Large-scale administrations of the instrument have indicated that the tasks are appropriate for second through fourth graders; appropriateness to other grade levels remains to be determined. The responses of several thousand children to the tasks have been overwhelmingly positive. Be they from open or traditional classrooms, virtually all have fun in the measurement situation.

PLAT procedures are intended to tap two molar dimensions of written language production–linguistic competence and communicative competence. Linguistic competency variables measure formal-mechanical aspects of the language that children produce independent of its content and functional or interactional qualities. They are measures of language *qua* language and represent the emphases of traditional education and educational evaluation. The linguistic competence variables incorporated in the PLAT, however, are derived from samples of connected discourse elicited in situations fostering divergent behavior rather than from multiple-choice responses to convergent questions as in traditional achievement tests. They include quantitative counts of fluency (length of the story); syntactic maturity (average length of T-unit or single independent predication together with any subordinate clauses or phrases that may be grammatically related to it); and vocabulary diversity (type-token ratio).

Communicative competency variables measure features of content and the functional quality of written language produced for purposes of communicating. They are measures of the success and sophistication with which language is used to convey meaning. Traditional educational evaluation has made virtually no attempt to tap these dimensions of language behavior. Communicative variables at present include counts of descriptive words and constructions and of explanatory statements and ratings of reporting quality and narrative organization. Some of the variables described here are currently being modified or deleted and new variables are being added.

Mutual Problem-Solving Task (MPST). The Mutual Problem-Solving Task (MPST) was developed to measure the potential long-term effects of participation in the Ypsilanti-Carnegie Infant Education Project on the relationships between mothers and their children (see Epstein, 1976). The measurement situation was designed to be closely analogous to situations that occur naturally in the home. Although the MPST measures both verbal and nonverbal behaviors, the description here focuses upon the assessment of children's language production.

Mothers and children are observed baking cookies together in their own homes. (Other culturally appropriate cooking activities could be used elsewhere.) Families are told there is no right or wrong way for them to act and that "we are just interested in learning more about how mothers and children work together." Mothers and children are given a choice of recipes, utensils, and ingredients so that decisions have to be made and their relative involvement in the decision making process can be assessed. A trained observer categorizes the behavior of mother and child as they prepare the cookies for baking and audio recordings also are made.

The MPST has been administered to approximately fifty first and second grade children and their mothers.

The task takes an average of twenty minutes and has proven enjoyable for both mothers and children. The observation method seems applicable to a broader age range of children, to different problem-solving groups (e.g., two children, teacher and child), and to different problem-solving tasks.

Observations are made using the Interaction Category System, a continuously coded observation schedule in which behavior is recorded sequentially. Interactional variables represent measures of communicative competence in an interactive problem-solving setting. They include measures of conversational reciprocity, reliance upon verbal communication and effectiveness in requesting information. Audio tape recordings are analyzed later by applying a modified version of the PLAT analysis procedures.

Derivative Instrument Development Efforts. Success with the PLAT and MPST has inspired three related efforts. First, efforts are being made to develop an oral version of the PLAT in which oral presentations (sustained discourse on a topic rather than conversation) are elicited from children. Thus far, it seems easier to elicit make-believe stories than reports. Second, various procedures for eliciting and analyzing the oral language production of preschool children are being developed. It has been substantially more difficult to create measurement situations for preschoolers which elicit sustained connected discourse than for elementary-age children. The third effort involved adapting the MPST for use with elementary-age children. Two measurement situations are being devised. In both, children work in small groups to solve problems involving real objects—first, a prestructured problem and next, a problem of their own design. Live observations will be made of verbal and nonverbal interactions during the tasks using a modified version of the MPST Interaction Category System.

General considerations of validity and reliability. Serious attention has been paid to instrument validity and reliability. Situational variation in the PLAT (oral and written) and its preschool version is low and on a par with other standardized assessment procedures. Situational variation in the MPST, however, is relatively high due to variation in physical aspects of the home environment and in maternal responses. Situational variation in the adaptation of the MPST for elementary-age children is somewhat less as a result of controls exercised over the physical environment. But does variation in the measurement situations necessarily invalidate the MPST and its adaptation? If one is trying to ascertain the productive language ability of individual children, the answer must be "yes."

If, however, one is evaluating the effects of an educational program on children's language production by comparing the mean performance of children participating in the program with that of like children not participating, *the answer may be "no."* If, for example, variation in the measurement situation is equally distributed across groups being compared, situational variation does not impair the validity of either the instrument or the comparison. Moreover, if the educational program is designed to produce certain outcomes in the language produced by children in the context of their relationships with their mothers at home, these outcomes can only be measured within the context of the criterion situations with all of their inherent variability. Of course, a successful program is likely to alter home context as well as the child's behavior. Finally, participants in the problem-solving task may be eliminated as sources of situational variability influencing outcome behaviors if the diad or group is treated as the subject of analysis rather than an individual. In the MPST, this can be accomplished by measuring interactions rather than the behavior of either mother or child.

Historically, psychometrics has focused on the problems of assessing individual behavior. The concerns and needs of educational program evaluation are quite different and appear to offer many opportunities for developing alternative assessment methodologies without the psychometric constraints associated with individual assessment.

MEASURING COMMUNICATIVE COMPETENCE

S.J.Savignon

Second language teaching today in the United States may be characterized as essentially "audiolingual" in methodology. The audiolingual method derives from the assumption that language is a set of habits which can be described, practiced, and measured. Its goal is communication, and the pattern dialogue with its emphasis on contemporary, idiomatic language use marked a welcome break with the grammar-translation method which looked to literary masterpieces for its models.

The fact is, however, that spontaneous communication rarely occurs in audiolingual classrooms. Manipulation of a carefully sequenced set of linguistic patterns has not proven to be the key to second language (L2) development. Moreover, the insistence on memorization, repetition, and avoidance of errors discourages development of strategies necessary for successful spontaneous interaction and thereby successful learning. The audiolingual method is no longer reflective of current thinking in the fields of either psychology or linguistics, and the current need is to develop teaching and testing strategies which meet *functional* goals of L2 use (e.g., Jacobovits, 1970).

As an example of one such teaching strategy (Savignon, 1972a), in one intermediate college French course, short wave news broadcasts from France were tape recorded daily and available to students at any time by dialing a telephone number. In class, students reported regularly on the preceding day's news, stimulating others to supply additional and even conflicting information. As communicative functions are defined in teaching, so must evaluations of L2 proficiency measure language use in *real life situations*. Discrete-point measures of competence in terms of the elements of language apart from an act of communication are not valid measures of functional skills.

Savignon (1972b) demonstrated experimentally the distinction between discrete-point tests of linguistic competence and tests of communicative competence. The research involved three groups of beginning college French students enrolled in an audiolingual program. For one group, a program designed to encourage the development of communicative skills was substituted for the usual hour of language laboratory drill. Occasions for meaningful use in these sessions included impromptu role playing, games, and discussions on topics of the students' choice. Contrary to recommended audiolingual practice, these students were encouraged to take risks, to go beyond what had been introduced in their regular class in order to express their own meanings. Errors were expected. The teacher served as native informant rather than drillmaster or judge.

At the conclusion of the eighteen-week course, specially developed tests of communicative competence were administered to students in all three groups, in addition to standard measures of achievement. While the latter measured proficiency in terms of ability to manipulate patterns (phonemic, syntactical, or lexical), the former defined a set of *occasions* for language use. Evaluation of performance was based on criteria set by a group of native speakers who were *not language teachers*. As nonprofessionals, so to speak, they would presumably not be accustomed to analyzing language in terms of separate elements but would respond to it functionally, in terms of meaning conveyed. The students were told these were tests of how well they could communicate in French in a variety of situations and that evaluation would be based on how well they got their meaning across. They were to concentrate, therefore, not so much on speaking perfect French as on using every means at their disposal to express their ideas and make themselves understood.

Four different communicative contexts were included:

1. **Discussion**—A four-minute informal interaction between student and native speaker to exchange as much information as possible on an assigned topic. The native speaker rated the student's effort to communicate and amount of communication.
2. **Information-getting**—A four-minute interview in which the student was to learn as much as possible about the person with whom he had conversed, take notes, and write a report in English. Evaluative criteria were: comprehensibility and suitability of the introduction; naturalness and poise; comprehension by the native speaker; comprehensibility and suitability of the conclusion; and amount of information obtained. The latter was measured by counting the correct statements in the student's write-up; the native speaker rated performance on the other criteria.
3. **Reporting**—A three-minute report on an assigned topic in which students spoke first in English to organize their ideas and then in French.
4. **Description**—A task to test the student's ability to describe an ongoing activity. After seeing an actor perform various actions, the student described the actor and his activities in English and in French. A native speaker rated this and the preceding task from tape recordings on the basis of fluency (effort) and comprehensibility. The native speaker also wrote in French what he understood from the recordings, and another evaluator then scored these accounts for amount of information conveyed. In none of the tests was there a penalty for linguistic errors that did not affect meaning.

The results demonstrated the overwhelming superiority on tests of communicative competence of those students who had used French creatively throughout the course. These same students, moreover, performed as well as the other two groups on traditional tests of linguistic competence. The discrepancy between linguistic and communicative competence showed up most clearly in the reactions of students in the nonexperimental groups to tests requiring them for the first time to use the language they had been studying in a variety of real life encounters with a native speaker: "If this is an easy test, I just found out that I couldn't talk my way out of the airport if I flew to France." "Until this evening I was never forced to say anything except answers to questions or substitute phrases."

Tests which measure not *knowledge about* language but an ability to *use* language effectively in an exchange with a native speaker are by definition context-specific. They must reflect the needs and goals of the learner for the L2 functions he will be required to meet. The assumption underlying the discrete-point approach to testing language proficiency has been that testing linguistic elements separately affords a more "objective" evaluation than is possible in an admittedly subjective evaluation of performance in an integrated skill. Laudable as these efforts have been, they have failed to take into account the complexity of the communicative setting. In their emphasis on linguistic accuracy, moreover, they have served to discourage the development of the strategies which are necessary for the development of communicative competence. (For illustration of a child's strategies, see Savignon, 1974).

Observations in the Classroom

An alternative solution to the problem of ecological validity is to assess children's language in the setting of ongoing classroom life. In the language domain perhaps more than anywhere else, Eisner's (1969) concept of "expressive objectives" applies. We want children to have certain experiences as speakers and listeners both because we think these experiences are good in themselves and because we think they are a good medium for other learning. If we could document that these experiences occur, wouldn't that be enough?

Scriven's (n.d.) independent summative evaluation of the "child lore" program, *Pass It On,* developed at the Southwest Educational Development Laboratory by folklorist Richard Baumann, is an example of just such an effort. *Pass It On* consists of films, chants, jump-rope activities, riddles, and trading-time activities—"indigenous language arts" as Scriven calls it. There also are systemic features such as the combining of two grades for the program and the use of student aides from higher grades as group leaders. Scriven did no testing of any children. Instead, in multiple observations by more than one observer at a time, his staff compared specific features of classroom life while the program was in use with these same classrooms earlier and later the same day and with other classrooms where the program was not used.

Observers used a detailed five-page checklist to note, for example, how students used the materials, general student attitude toward materials, race and sex differences in use of the materials, immediate interpersonal effects, and speculations about long-term effects. With the exception of one program-specific section, the same checklist was used in all classrooms, whether or not using the program. Evidently, the observers made extensive annotations in addition to checks.

From these observations, without frequency counts or numerical ratings, Scriven can make such comments as the following:

> The overall judgment of the program is strongly favorable...an impression that was shared by every observer, operating independently and after discussion and reporting on sixteen different sites (p.9).
>
> [In the jump rope activities] we saw perhaps the most noticeable tendency toward producing social change in the children's behaviors. There was a widespread antipathy towards jump rope games by the boys when they were first introduced, but it was an antipathy that rapidly evaporated and almost universally turned into a highly participatory and enjoyable experience. [The handclap games] were also very successful, although—it was our impression—it was for a rather smaller number of the students. This is as good a time as any to stress the very successful integration of kinesthetic with cognitive and affective dimensions in this program (p.12).
>
> It's time to turn to the less good news. The kindergarteners don't get the riddles at all...The "trading-time" activities were not really successful, at least in the more obvious dimensions.... There are two rather more serious implications than at first sight appear. First, there is...the possibility of some tension between the relatively structured and authoritarian presentation times in the first part of the week and the expectation of free spontaneous activities in trading time. The other, graver, implication can best be expressed in terms of a discussion that I had with one boy [about how he had been first to raise his hand and then been thrown out for telling a "Polish joke"]. The teacher was neither prepared nor able spontaneously to handle this intrusion of racism into riddling. There were consistently no racial distinctions except those made by teachers in asking possibly more black children than white to exhibit stylish chanting techniques (pp.13,15,25).

In comparison with these observations of Pass It On, observers found only "minimum necessary" attention in their nontreatment observations and an almost total lack of student interaction.

Scriven's evaluation design was "goal free." Only after the observations were written up did he and his staff consult the manual and compare what they had seen with what the program designers intended. On this basis, they question whether the materials achieve "experience in creativeness" or enhance linguistic and communication skills.

There are two limitations to relying totally on observations of naturally occurring classroom life. One, discussed by Shuy in his paper, is the "How do you know it didn't happen when the camera was off?" problem. The other is that no natural situation elicits equally from all participants. Individual children in any classroom do not contribute either equally or randomly to the interaction which occurs. Particularly in those more "open" educational programs that are most interested in communicatively-based assessment of oral language, children construct their educational environments in individually different ways (Stodolsky, 1972). In such environments, it is not enough to know that productive talk and language experiences of desired kinds are "in the air." Even if one wants only an evaluation of the environment, not diagnostic information on individual children, one still must somehow monitor the participation of individual children—either all of them or some randomly selected sample. Otherwise, we may be overlooking the language of the more silent children, some of whom are the ones the program should be designed most to help.

With this in mind, assessing any environment for language use by naturalistic observations assumes large and usually unfeasible dimensions. The density of evidence on whether, for example, particular children ask questions as well as answer them or construct coherent narratives or explanations is likely to be thin, even over the course of extended observations. So one must seek supplements to naturalistic observations in situations which can yield more information in less time. These supplements then merge with the concentrated encounters described above.

Focal Aspects of Communicative Competence

What is important about both the High/Scope and Savignon work is not just what they have done but what they are striving for. Both are attempts to assess the outcomes of curricula which are themselves based on true intentional communication, and the assessment situations are concentrated versions of that ongoing classroom life. In both, competence can be defined as high quality performance in important life situations, and assessment of that competence is in terms of functional effectiveness rather than formal correctness. Furthermore, both High/Scope's mother-child problem-solving task and Savignon's communicative tasks permit evaluation of communicative strategies when the speaker's language repertoire is inadequate for the task: how the child supplements words with gestures to achieve his goal; how the foreign language learner asks for help with words not known.

But how do we decide which communicative functions are of the most worth? One can not fault High/Scope or Savignon's objectives–reporting, narrating, describing etc. But where does the list of communicative competencies end? Can we establish some principled alternative to an unclustered list–surely the least useful cognitive framework for teachers and evaluators alike?

Two California school districts provide maximal contrast on this question. District A has a "language continuum" on which each classroom teacher checks the appropriate skills observed in the classroom performance of each student. The continuum has twenty-one receptive items (from "Points in direction of the source of a sound" through "Interprets material through dramatic play, role playing, or pantomime") and fifty productive items (from "Expresses needs and wants verbally" through "Gives oral reports"). District B focuses on "speaking relevantly." Each teacher conducts an activity such as a class meeting or creative dramatics while the aide observes and records. During the first ten minutes, all of each child's responses are rated as relevant, irrelevant, nonparticipant, or goofing off.

A list of fifty or a focus of one? Interestingly, in research at the Center for Applied Linguistics (CAL), one aspect of children's functional language competence which appears to differentiate between the children considered more and less competent by their classroom teachers is what Peg Griffin (personal communication) calls "speaking topically." Perhaps some intersection of theoretical work in pragmatics (e.g., Searle, 1976) and empirical research with children as at CAL will suggest where to focus.

But even then, problems will remain about evaluation criteria. "Speaking relevantly" is a particularly interesting case in point. Students in my university classes have pointed out a contrast between ways of entering class discussions that are differentially responded to by some professors. In Philips' (1974) sense, contributions to class discussion based on narratives of personal experience don't "get the floor." And in heated discussions around a conference table, narratives based on personal experience are sometimes dismissed as no more valid than testimonials at a revival meeting. Are the narratives relevant? Who is to decide?

Apart from increasing successful execution of particular communication competencies, an educational program can increase the range of contexts in which such competencies are manifest. Conceptually, it is probably wrong to talk about the complete decontextualization of any skill. Transfer is always to some place; none of us can speak equally effectively in any and all circumstances. One important purpose of education is to increase the range of such circumstances for each child.

To assess this educational effect, we have to construct encounters which sample a range of situations on some principled basis. Vulpe, Rollins, and others (Vulpe, in press) have made this aspect of growth a central component of an assessment system for children with special needs. Their "performance analysis scale" is based on the concept of "engagement" as an expression of the child/environment interaction. Rather than only looking at behaviors the child demonstrates under fixed conditions, it looks at the ways in which changes in the social or physical environment may affect changes in the child's pattern of response. While this scale is not primarily about language and is designed especially for children under 3, the underlying ideas have more general application.

Quantitative and Qualitative Evidence

So far I have said nothing about the use of numbers; questions about "how" to assess have to follow discussion of "where" and "what." To some extent, the use of numbers will change if assessment procedures shift to the kinds discussed above. Test scores consisting of number of items correct will be less used; ratings of functional effectiveness will increase; and mean length of utterance (MLU) will eventually be replaced. Since MLU, or variations on it, is still so widely used, it merits further discussion.

In language development research, a traditional measure of children's language competence is obtained by computing the mean length of 50 or 100 utterances. In research with school age children, an adaptation of MLU is used. Instead of an utterance, a "minimal terminal unit" (T-unit) is substituted. A T-unit is one main clause with all subordinate clauses attached to it.

Problems about MLU (or T-unit) involve its validity, reliability, and informativeness.

The validity of MLU as a measure of young children's competence rests on the widespread finding that it increases with age and on discoveries of a correspondence between mean length and the appearance of specific grammatical features (such as auxiliaries). One trouble with length as an indicator of complexity, however, is that one counts units (words or morphemes) as if they were beads on a string and all alike. This is true of numbers in a digit-span test, but it is not true of words in normal sentences. A sentence is special precisely because it has an internal structure, and the units in that structure must have differential cognitive weight.

As a measure of spontaneous speech or writing samples, MLU also has severe problems of reliability. Even with the very young child, MLU will vary with the pattern of conversation—for example, with the density of one-word answers to questions. Situational influences assume almost prohibitive importance beyond the preschool years.

Finally, there is the question of informativeness. If two children or groups of children differ on MLU, we can rank them accordingly. But what else have we learned? We can say nothing about what they do or do not do with language. When MLU increases in older children as a result of particular educational experiences, it is hard to assert what has been learned. Numbers alone cannot tell us what has happened. In current terminology, MLU is a norm-referenced rather than a content-referenced measure and, as such, deserves to be replaced in evaluation research.

An important question about the proper role of numbers is how to combine productively quantitative and qualitative descriptions. Since most assessment now is quantitative and the availability of computers is itself a press toward the use of numbers, it is especially important to consider this limitation and the essential information that qualitative evidence can contribute. Sieber (1973) has a full analysis of "the integration of fieldwork and survey methods" in sociological research. Throughout language evaluation research, it would be useful to seek such combinations, or the unanswered questions (see Cazden, 1976, and in press,b) that remain without them. Here I will focus on only one contrast: the need for qualitative data when the purpose of assessment shifts from summary description to process explanation, from program evaluation to diagnosis of either children or their environments, and to hypotheses about points of leverage for change.

In most of the above examples of "what" to evaluate, primary focus is on language function, on the speaker's integration of linguistic elements in an intentional communicative act. That is, I think, as it should be. Yet for individual diagnosis, it becomes necessary to compare more and less functionally effective communications in terms of the discrete elements which are used or omitted. In general, functional need stimulates the development of formal resources to meet those needs. But not always so. And where it does not, the focus of attention of diagnostician or teacher (and maybe even temporarily of the learner) must shift to form. In a discussion of "the form/function paradox," Crystal (1975:40-41) criticizes for failing to understand this point.

> ...Given two children engaging in a "use" of language, how is one to judge their relative success, or influence the less successful to improve? Apart from any pedagogical problems, the teacher must carry out at least four preliminary tasks: (1) identify the differences between the two, i.e. determine which features of language account for their differing performance... (2) he must be clear as to the salient linguistic characteristics of a "good" example of the language use being aimed for; (3) once he has made a diagnosis, he must be aware of the possible linguistic pathways towards achieving this use of language...and (4) having decided to implement a particular line of action, he needs to be able to identify progress—which amongst other things involves an ability to identify unexpected or misleading linguistic developments, such as the emergency of a structure which in fact militates against the development of the target use of language.

When such attention to linguistic structure is necessary, quantitative measures such as mean length of utterance (or T-unit) cannot suffice. Summary statistics must be informed with more qualitative detail and concrete insight—in this case from linguistics—into the patterns of speaking from which the statistics come.

The same is true when the diagnosis is not of children but of their environments—a critically important focus for more evaluation research. Consider the observations of teacher-student interaction conducted on behalf of the U.S. Commission on Civil Rights in 494 elementary and secondary school classrooms in California, New Mexico, and Texas (reported in Jackson and Cosca, 1974). The report is a damning document:

> Teachers praised or encouraged Anglos 35% more than they did Chicanos, accepted or used Anglo's ideas 40% more than they did those of Chicanos, and directed 21% more questions to Anglos than to Chicanos. Thus, Chicanos in the Southwest receive substantially less of those types of teacher behavior presently known to be most strongly related to gains in student achievement (p.227).

And this in classrooms which had been selected from only those schools with no previous record of civil rights violations or investigations and in which teachers were aware that an observer from a federal civil rights agency was present.

While such quantitative evidence may be a sufficient base for legislative action or legal decisions, it is not sufficient to guide attempts at change. When someone tries to move from condemning such environments to planning more culturally responsive education, more detailed qualitative analyses are necessary—here sociolinguistic and ethnographic—of patterns of speaking in classroom and outside.

In the conclusion of her ethnographic study of sociolinguistic interference in the classrooms on the Warm Springs Reservation, Philips (1974: 311-12) comments on the 1974 U.S. Commission on Civil Rights report:

> The orientation of the Commission report is such that cultural differences...are not dealt with in attempting to account for the disparities discussed. The impression is given that the disparities are due to what is typically referred to as "discrimination." But...even where teachers are well intentioned, the results are similar, because the minority students' efforts to communicate are often incomprehensible to the teacher and cannot be assimilated into the framework within which she operates. The teacher, then, must be seen as uncomprehending, just as the students are. And it is primarily by virtue of her position and her authority that the students and not the teacher come to be defined as the ones who do not understand.

As a non-ethnographer, I am sure that only qualitative analyses can illuminate how such misunderstanding is produced in actual classroom events. Numbers alone cannot explain either how those numbers came to be or what can change them.

References

Bond, J. T. "The High/Scope Productive Language Assessment Tasks: Effects of the Cognitively Oriented Curriculum on Follow Through Children's Written Language Production." Ypsilanti (Mich.): High/Scope Educational Research Foundation, 1975.

Bronfenbrenner, U. "The Experimental Ecology of Human Development." (Mimeo.). Ithaca: Cornell University, n.d.

Cazden, C.B. "Hypercorrection in Test Responses." *Theory Into Practice* 14:343-46, 1975.

Cazden, C.B. "Review of Cicourel, A.V., et al., *Language Use and School* Performance." *Contemporary Psychology* 21:125-26, 1976.

Cazden, C.B. "How knowledge About Language Helps the Classroom Teacher, Or Does It? A Personal Account." *The Urban Review,* in press(a).

Cazden, C.B. "Review of van der Geest, T., Gerstel, R., Appel, R. and Tervoort, B., *The Child's Communicative Competence: Language Capacity in Three Groups of Children from Different Social Classes.*" *Language in Society,* in press(b).

Crystal, D. "Linguistic Perspectives." *Reading* 2:37-49, 1975.

Eisner, E.W., "Instructional and Expressive Objectives: Their Formulation and Use in Curriculum." In *Instructional Objectives.* AERA Monograph No. 3. Chicago: Rand McNally, 1969.

Epstein, A.S., P. Schwartz, J. Meece, D. Lambie, P. Dukes, F. Crawford, and P. Phillips. "The Mutual Problem-Solving Task." Ypsilanti (Mich.): High/Scope Educational Research Foundation, 1976.

Gagné, R.M. "Observing the Effects of Learning." *Educational Psychologist* 11:144-57, 1975.

Jackson, G., and C. Cosca. "The Inequality of Educational Opportunity in the Southwest: An Observational Study of Ethnically Mixed Classrooms." *American Educational Research Journal* 11:219-29, 1974.

Jacobovits, L.A. *Foreign Language Teaching.* Rowley (Mass.): Newbury House, 1970.

Philips, S. "The Invisible Culture: Communication in the Classroom and Community on the Warm Springs Indian Reservation." Unpublished doctoral dissertation. University of Pennsylvania, 1974.

Savignon, S.J. "A l'écoute de France-Inter: The Use of Radio in a Student-Centered Oral French Class." *French Review* 46:342-49, 1972(a).

Savignon, S.J. *Communicative Competence: An Experiment in Foreign Language Teaching.* Montreal: Marcel Didier, 1972(b).

Savignon, S.J. "Talking with My Son: An Example of Communicative Competence." In F. Grittner (ed.). *Carrers, Communication and Culture.* Skokie (Ill.): National Textbook, 1974.

Scriven, M. "Evaluation of Southwest Education Development Laboratory's Children's Folklore Project: 'Pass it On'." Berkeley (Ca.); Educational and Development Group, n.d.

Searle, J. "The Classification of Illocutionary Acts." *Language in Society* 5:1-23, 1976.

Shapiro, E. "Educational Evaluation: Rethinking the Criteria of Competence." *School Review* 81:523-49, 1973.

Sieber, S.D. "The Integration of Fieldwork and Survey Methods." *American Journal of Sociology* 78:1335-59, 1973.

Stodolsky, S. "Defining Treatment and Outcome in Early Childhood Education." In H.J. Wahlberg and A.T. Kopan (eds.). *Rethinking Urban Education.* San Francisco: Jossey-Bass, 1972.

Vulpe, S.G. *The Vulpe Assessment Battery, Developmental Assessment, Performance Analysis, Program Planning for Atypically Developing Children.* National Association for Mental Retardation, York University Campus, 4700 Keele St., Toronto, Ontario, Canada, in press.

CRITIQUE

Dell Hymes
University of Pennsylvania

Rather than responding point by point to the papers presented by Shuy and Cazden, I should like to offer a more oblique form of critique in which I shall refer to their papers while attending both to the central theme of the conference, the question of qualitative and quantitative research methodologies, and more specifically to the uses of linguistic ethnography in assessing language development. In doing so, I shall briefly comment on the development of linguistic methodology before turning to the uses of linguistic ethnography in education.

Linguistic methodology offers two important lessons in coming to terms with the relations between quantitative and qualitative methods. The first lesson derives from the success of linguists in this century in discovering relationships in a central aspect of human life that are capable of rigorous formulation, of patent reliability and validity, without recourse to numbers. Linguists such as Sapir and Bloomfield created a qualitative methodology which, on the one hand, generalized insights into particular patterns of speech sound from study of particular languages and, on the other hand, transcended the phonetics of pure physical measurement. In effect, they replaced rigor of measurement with rigor of functional contrast.

Sapir's essential point was the distinction between a physical event and an element in a system of signs. Two languages might have identical inventories of sounds according to observation of physical properties; yet when the functional relations among the sound within the system of the language were considered, the two languages might have quite different patterns, configurations, or structures of elements. The difference would lie not in the presence or absence of observed sounds but in the status of the observed sounds within the system of the language. And the principle that determines the status is qualitative, all or nothing, leading to invariant, fixed reference points. There is rigor in such work and a branch of scientific inquiry to which to appeal, but it is qualitative and discrete mathematics, algebra, logic—not statistics or experimental measurement.

The researcher armed with qualitative methodology can be just as a priori in assumption, just as prone to overlook disquieting empirical facts, just as heavy handed in the service of his metholodogical god as can the quantitiative research of fabled evil, But the lesson remains. Any consideration of qualitative methodology in the study of human life must take into account the success of linguistics in establishing a sector of study that has a methodology that is at once qualitiative and rigorous.

Whereas the first lesson has to do with validity in the sense of structure, the second has to do with validity in the sense of function. Sapir showed in regard to phonology that recognition of structure depends upon recognition of functional relevance, that the key is not the relationship of sound to sound alone but in the service of distinguishing units of another kind (words, sentences). Linguists have repressed and learned this lesson of functional relevance again and again—for phonology (as against phonetics), for morphology, for syntax, and for semantic relationships.

Each functional sector or level of language organizes units in a way not given by the units themselves. To be sure, as Cazden points out quoting Crystal, one must attend to the specific units of language or one will not see any relationships at all (just as ignorance of the speech sounds of a foreign language will yield a sense of noise, not phonology). But the relationships that are there will not all come into view if one stays at the given level. One must start from the functional category and discover what elements and relationships among elements may serve it. Shuy's studies of functional language illustrate this principle in their examples of alternative ways to accomplish requests, directions, instructions, and the like.

This brings us to the last leap in applying the principle of functional relevance, to the study of the relationships among linguistic elements in the service of speech styles. Here we are concerned not with another set of units parallel to phonemes, morphemes, syntactic constructions, and semantic elements, but with a novel organization of all these units. The defining attributes of a style may differ quantitatively from the defining attributes of levels such as phonology and syntax. While some differences among styles may depend upon all-or-nothing contrasts, others (as Shuy illustrates) depend upon proportions and frequencies that appear only when one sets out to discover them from social life, not from grammar.

What we have come to, then, is a study of language that is inescapably sensitive to situation and in which quantitative differences are inseparable from qualitative effects. As the papers by Cazden and Shuy show, such a study of language is beginning to emerge into prominence, and it is the study of language that is fundamental to education. Too few are engaged in such study—the price we pay for the isolation of linguists and educational research from each other, for disabling polarizations between qualitative and quantitative methodologies, for the lack of a cadre of linguistic ethnographers.

We cannot adequately evaluate language development and the uses of language that enter into education without attention to the principle of contrastive relevance—to the demonstration of functional relevance through contrast, showing that a particular change or choice counts as a difference within the frame of reference. Properly pursued, the extension of the principle of contrastive relevance in linguistic ethnography entails a conception of language development and use as a matter of meaningful devices. The still common use of mean length of utterance as a measure of development, to which Cazden alludes, is not consistent with this principle. While the measure may helpfully correlate with other things, it can shed no light on what is happening, on what is being acquired and used.

Language, from sound to style, is a complex of form-meaning convariation. That is another way of putting the point of contrastive relevance. To discover what is there, what is happening, one seeks to discover what changes of form have consequences for meaning and what choices of meaning lead to changes of form. One works back and forth between form and meaning in practice to discover the individual devices and the codes of which they are part.

The limitation of linguistics proper for the study of language development is that it tends to stop short of the full range of form-meaning convariation and to stop short of ethnography. Linguistics elaborates discovery of the fact that a feature is meaning relevant, has structural status in terms of a function, and it works readily with first order approximations of meaning content—the glosses that are immediately available. It tends not to pursue meaning in terms of resonance and consequence.

The principle of the linguistic ethnography that is needed can be put in terms of complementary perspectives. If one starts from social life in one's study, then the *linguistic aspect of the ethnography* requires one to ask: What are the communicative means, verbal and other, by which this bit of social life is conducted and interpreted? What is their mode of organization from the standpoint of verbal repertoires or codes? Can one speak of appropriate and inappropriate, better and worse, uses of these means? How are the skills entailed by the means acquired, and to whom are they accessible?

These questions lead into the territory of the other starting point. If one starts from language in one's study, the *ethnography of the linguistic work* requires one to ask: Who employs these verbal means, to what ends, when and where and how? What organization do they have from the standpoint of the patterns of social life?

We must also, I believe, consider our own uses of language as scholars and scientists. To the best of my knowledge, some of what we learn and should convey can be expressed only through skillful prose. In anthropology and in personal life, much of what we know is known through narratives, anecdotes, firsthand reports, telling observations. But in our scholarly chairs, we find it difficult to acknowledge their validity. If we are to extend our understanding of language to the full, so that we can fully comprehend its role in schooling, in education, in social life, in our own lives, we have to find a way to come to terms with the validity of uses of language that are aesthetic. Indeed, such uses do play a vital part in decisions and perceptions, so that we handicap our understanding of educational institutions and the forces that affect them if we do not make them explicit objects of attention. Our own language development is in need of assessment.

CRITIQUE

Blandina Cardenas
Center for the Management of Innovation
in Multicultural Education
San Antonio, Texas

Most of my professional life has been spent in advocating, developing, and implementing education strategies designed to equalize educational opportunities for culturally different children in general and for Mexican-American children in particular. It is from this perspective that I respond to the papers by Shuy and Cazden.

Both Shuy and Cazden have approached the issue of assessing language development in ways that make me eat my past words about linguists. It's not that prejudice against linguists is something I learned at my father's knee. It's just that as a practitioner I have found myself face to face on a daily basis with teachers who have endured countless inservice sessions on how to make Mexican-Americans stop saying "sheep" for "cheap" and "sheet" for "cheat." In fact, most proposals for preparing bilingual teachers have as an overriding goal the development of contrastive analysis skills in teachers. Obviously, the gap between the ideas underlying the research reported by Shuy and Cazden and those underlying most bilingual education programs is very, very great.

The policy decisions that have given impetus to biligual education—in Lau vs. Nichols and other court cases and in various bilingual education acts—have proceeded from the view that language is the critical factor in the denial of equal educational opportunity to culturallly and linguistically different children. In Lau vs. Nichols, for example, involving Chinese children in the San Francisco schools, the Supreme Court ruled unanimously that a student who was given the same textbooks, classroom, teacher, and so forth as other students was not being provided equal educational opportunity when he came to school with a different language. Perhaps if we substituted functional communication for language we would be closer to the truth in identifying the critical factor, but we know that language is but one of the elements that impairs equal educational opportunities for the culturally and linguistically different.

Shuy observed in his paper that the policy decisions underlying the establishment of bilingual education programs have preceded the research with which to implement them. While I agree that we do not have an adequate research base for designing and assessing programs for bilingual education, I would not agree if someone were to infer from Shuy's statement that we should have engaged in extensive linguistic research prior to the enunciation of policy. To have done so could well have delayed the establishment of policy while research was conducted on only one aspect of the total problem. Moreover, it is important to remember that the Court's decision in Lau vs. Nichols did not require the establishment of bilingual education programs. The plaintiffs stated, "We do not seek the enunciation of a remedy around these issues," and the Supreme Court did not, in fact, stipulate a remedy.

All of this is to say that bilingual education programs should not be assessed from a narrow linguistic perspective. My own prejudice is that the exclusive use of either psychological or linguistic perspectives in assessing bilingual education programs would not serve the society, educational decision makers, public policy makers, or the children who are supposed to benefit from the programs. Bilingual education must be viewed as a complex national phenomenon which has attached to it deep-seated attitudes about language, race, the national identity, economics, and class. It should therefore be studied from a sociological, anthropological, and political view, as well as psychologically and linguistically. Assessment of bilingual education will be futile if we do not face up to the complex and fundamental social issues which cut across and color all that we do.

In my own part of the country, for example, bilingual education has emerged as a symbol for a whole series of aspirations about political, economic, and social equality. It has become the vehicle by which minority group educators have earned some place in the total decision-making process about schools. In assessing bilingual education, then, we need to understand how it acts as an innovation within school settings and how schools act upon it. We need to differentiate among issues of evaluation that are related to all of the deep-seated attitudes and aspirations attached to bilingual education and those that are related to the innovation itself—the evaluation issues peculiar to bilingual education as a pedagogical strategy and to the process of innovation in general.

We will need both quantitative and qualitative data for these assessments. We need quantitative research because policy makers will be looking for that type of data to reinforce a predisposition to support the concept of bilingual education. And we need qualitative data because at the school district level, where much of the work in bilingual education is being done, qualitative research information helps teachers, administrators, and community members to make the leap from quantitative data to their intuitive grasp of classroom reality. Both types of research are necessary since they appeal to different audiences, provide different perceptions, and lend themselves to somewhat different emphases in the investigation of specific issues.

Whether quantitative or qualitative, research should be looking at what happens in classrooms rather than perpetuating the practice of looking at what's wrong with children—linguistically, developmentally, socially. Cazden presents two promising efforts to provide alternatives to the traditional testing strategies—the assessment procedures used by High/Scope and by Savignon. I think these approaches hold particular promise for the evaluation of developmental efforts by R&D organizations and by a very few school districts involved in curriculum development. With time and perhaps some modification, the strategies Cazden presented might find wider use by school districts in implementing innovative programs.

It is my belief that the challenge of responding to issues of linguistic and cultural diversity is a permanent part of the American educational scene. Investigation of our responses to this challenge should occupy a more pervasive part of our educational research stream. Indeed, I expect that the assessment of bilingual education programs will become a highly visible, controversial, and expanding concern of educational R&D as legislative, judicial, and community pressures increase. I hope that we will conduct these assessments using both qualitative and quantitative research and recognizing that the issues involved go beyond language to embrace a complex of social, political, and economic attitudes and aspirations.

WHY DO DEMONSTRATIONS?

Growing out of the development of demonstration projects and the implementation of programs is the need to identify what, if anything, is different and what is changed as the result of installing such an innovation. How can such projects be evaluated, and what can we learn from such evaluation that can be applied to future implementation? What are the crucial elements to success, and how can these be identified?

WHY DO DEMONSTRATION PROJECTS?

John E. Dawson
Defense Resources Management Information Center
Naval Postgraduate School

Why do demonstration projects? Why, indeed! The question is timely, if not overdue, and even approaching an answer is complex. It is, however, a very subordinate question. Demonstration projects are just one variation in an alleged chain of actions that might bring different concepts to successful application throughout public education.

The larger, more basic, question is, Why do educational innovation? Rather than simply pricking my finger on the bramble of demonstration projects, I might as well bloody my entire anatomy by crashing the thicket called educational innovation. Incidentally, the question, Why do educational innovation? also draws the response, Why, indeed! That is fair warning that what follows is more attuned to inquiry than faith.

Demonstrations: What Are They?

The term "demonstration," according to Webster, suggests several possibilities: (1) "a course of reasoning showing that a certain result is a consequence of assumed premises"; (2) "an act of demonstrating, a means of proof"; (3) "an act of showing and emphasizing of the salient merits, utility, efficiency, etc.," of a product or service; (4)an outward expression or display, as of feelings"; a manifestation of emotions as by a crowd.

Acts labeled as demonstrations may reflect a muddling together of these concepts or utilize them in meaningful combination. For an example of the latter, consider demonstrations by Mercedes-Benz. Their engineers utilize a course of reasoning (1 above) in designing potential changes to the car and take test models to their famed test track as a means of proof (2) prior to introducing changes into mass production. The dealers are given demonstrators (3) with which to display salient features to potential customers—a process of dissemination and adoption. Finally, emotion-heightening film clips from the test track are shown on television in the hope of stimulating the crowd of watchers to see the demonstration at their nearest dealership. The last may be a rather perverse treatment of the emotional side (4) of "demonstrations," but it is one way of suggesting that demonstrations by organizations need not be barren of emotional content or purpose.

This brief illustration suggests that demonstrations may be of differing kinds and fulfill multiple purposes. It also suggests that any of these possible forms of demonstration is part of a chain of actions and means little in isolation. The test track and demonstrator cars have their significance in relation to change and performance within the total sphere of Mercedes-Benz. By analogy, educational demonstration projects have their significance in relation to educational innovation and performance within the total sphere of public education. We shall view them thus in our inquiry.

Approaching the Inquiry

Demonstration projects have great appeal to those wishing to influence the formulation of public policy. It is natural to feel that if we can only demonstrate on a small scale in our area of public service, we can thereby gain approval for large-scale application.

Demonstration projects also appeal to those of us interested in public policy analysis, public resource allocation, and public program performance. However, we approach them not as a tool of influence but rather with the analyst's typically doubting mind. Potentially, demonstrations may answer the analyst's three great questions: (1) Does the program/project/innovation/etc., do any good at all? (2) Does it do more good than it costs? (3) Is there another way to do more good for the same or less cost?[1] The sticky part of these questions is "good"—what kinds of good, for whom, to whom, by whom, when, where, how, and why! In order to examine demonstrations, innovations, and the larger thicket of public education, we have to find some "good." Our main thrust, then, is that we're looking for some gain by students. However, we will watch for other kinds of "good."

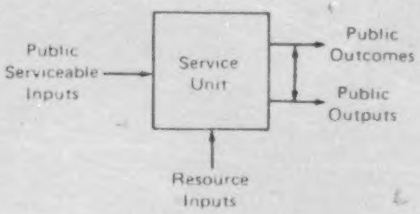
Along with most everyone else, analysts of public issues and programs today string their concepts together with models. While those we will use here have many elements in common with models already in disfavor in the educational innovation community, the distinctions to be made are more than subtle shadings, so try not to leap before you look. Our three models will be of the black box type: (1) the black box in a vacuum, (2) inside the black box, and (3) the black box in a context.

Beginning the Inquiry: Simplest Ingredients and Relationships

Figure 1 illustrates our first model—the black box in a vacuum for a public service. This is really a black box; note the lack of thruput identification (in arrows do not link to out arrows). Nevertheless, we do assume normally that at least one public outcome is some change to the public serviceable inputs while in contact with the service unit; there may also be other public outcomes. Resource inputs are bought by public funds (budgets) and used by the service unit. Things the service unit does which get out to the public are public outputs. Note that outputs and outcomes are not synonymous. Outputs are things done by government for (and often to) the public; outcomes are results external to government for, in and/or among the public. Public outputs and public outcomes usually are interdependent or interactive (Dawson, 1971).

FIGURE 1

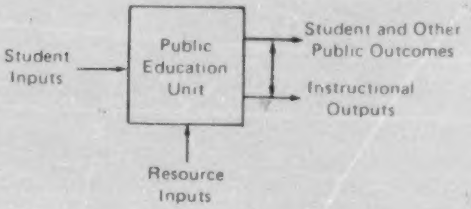
Model I
Black Box in a Vacuum—A Public Service



A tentative application of Figure 1 to public education is illustrated in Figure 2. At this stage, we normally would investigate (1) public education units (without invading the box), (2) main (student) outcomes and other outcomes, (3) outcome-instructional output relationships, (4) resource input-instructional output relationships, (5) resource input-student input relationships, (6) student input-student outcome relationships, and (7) other discovered relationships. In this instance, we will proceed through the first three steps and then stop for we shall have reached the central problem of educational innovation—ineffectiveness.

FIGURE 2

Model IA
Black Box in a Vacuum—Public Education



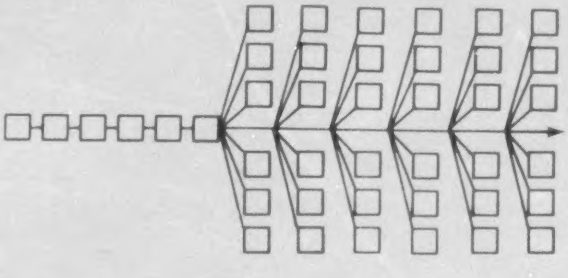
Public education units. Education is a field of human activity which is pre-industrial—a cottage industry. True, the little red schoolhouse is gone and there now are mega-universities. But the institutional structure of educational units is still basically a conglomerate of cottages, called classrooms, with a professional (or occasionally more than one) working at her or his craft in each cottage. To underline the obvious: the instructional outputs are dispensed by individual instructors. Thus, unlike Mercedes-Benz and its cars, an educational unit cannot display its demonstrator-model-teacher and expect students, parents, and community to agree that other craftsmen are essentially similar or that they are dispensing similar instructional outputs. The whole pattern of standardized, high technology, twentieth century production may be disfunctional when applied to guild members employed in cottages.

The public education unit or set of units forming a local educational agency operates a near monopoly for the community. While there are private schools, they do not significantly reduce the monopoly characteristics of local public education. The unit will have customers and will remain in operation regardless of the fate of an innovation.

From the viewpoint of students, a most important characteristic of educational service units is that they occur in series—in a series of classrooms and schools over time. Figure 3 provides a simple picture of one possible elementary-secondary series for a student. The

FIGURE 3

A Student's Series of
Educational Service Units (Classrooms)



series is connected only loosely. What may for the student represent "successful" student outcomes at one stage may not be appropriate preparation for the next stage. This is not simply a result of differences in the personalities of teachers or of the decentralized character of educational organizations. It is exacerbated by the fact that about 40 per cent of American families change addresses in five years. To the extent that innovation creates radically differing student experiences, it may greatly complicate the student's progression and total educational experience.

Outcomes. While the public education process is related to a wide range of public outcomes, attention here is limited to student outcomes; other outcomes will be noted in our third model. We shall consider two dimensions, behavioral and time, as presented in Table 1. Several comments should be made.

1. The timeline of analysis here is until government completes its output. For the Post Office, this timeline purportedly is measured in hours or days between the sending and receiving of a letter. In the case of public education, it is a minimum of about ten years to more than twenty years for the instruction delivered to a student.
2. There are intermediate outcomes and outputs in series during this long timeline. Examples of intermediate outcomes would be cognitive achievement in second and fifth grade mathematics and reading or even "successful" response to a specific fifteen-minute instructional unit. There are parallel intermediate instructional outputs.
3. The "goods" and one "bad" have been inserted with the intention of reflecting conventional wisdom, although each assignment of "good" or "bad" is arguable and consensus on any is illusive.
4. These outcome statements beg for a social environment—a context. (That is why black box in a vacuum models are so frequently found unsatisfactory when used as the sole tool of analysis.) However, contextual richness is premature and will await our third model.
5. The time section of Table 1 contains the far more explicit and quantifiable outcomes. They directly address the role of education in the life cycle of a people.
6. The behavioral section of Table 1 admits both cognitive and affective behavior outcomes. In the war between "tough-minded" and "tender-minded" evaluators (Kogan and Shyne, 1966), we shall not choose sides. In fact, both behavioral outcomes exist, whether intended or not.
7. Table 1 is neither exhaustive nor necessarily "right." Each of us is quite capable of making his own Table 1, and most such tables might well be better than this one. However, different perceptions of educational outcomes are a major part of the problem.

TABLE 1

Behavioral and Time Student Outcomes of Public Education

Behavioral

Some (good) change in cognitive and affective behavior of student inputs following termination of public education (ignores adult/continuing education)

Some (good) set of skills and expectations regarding work

Some (good) set of skills and expectations regarding social life

Some (good) set of skills and expectations regarding personal life

Time

Some (good) delay in entering work force

Some (good) shorter proportion of total life span involved in work

Some earlier entry into work force and proportionately longer work span in the case of dropouts (bad)

Output-outcome relationships. Next let us examine output-outcome relationships. Our original form of Model I (Figure 1) asserts a relationship of interdependence or interaction. My research within government over the last fifteen years indicates that this is nearly always the case. The exceptions are usually public programs that "pour concrete." For example, when a dam is built, stream fishermen abandon the stream (with much gnashing of teeth) and lake fishermen try out the new impoundment. The flow of action-reaction is one directional, at least in the specific instance. Or when a freeway is built, commuters adopt different routes and have different accidents, but the road does not adapt or iterate (although Shirley Highway outside Washington, D.C., is reputed to have been under continuous construction for the past thirty years). However, if the government output can be affected by its contacts with society, outcomes will feed back responses to service production and cause iterations and adaptations.

Thus, Model IA (Figure 2) hypothesizes an interactive relationship between outputs and outcomes in public education (and for intermediate outcomes and outputs as well). This relationship suggests that the teacher adapts to the students as they adapt to the teacher.

The nexus of instructional outputs and student outcomes is the focal point of effectiveness analysis. For purposes of decision about ongoing activities, our interest is in *change*, not the general *state* of the relationship. This is an important but often overlooked distinction and deserves a bit of explanation.

Instructional outputs are produced (teachers teach) and student outcomes occur (students learn). The process is "effective." The decision problem is: (1) whether a change in the controllable element (an innovation in the production of instructional outputs) induces a change in outcomes; (2) whether the change in outcomes is a net gain (i.e., involves movement toward value goals sought); and (3) if the answers to (1) and (2) are positive, how the degree of movement measured for an alternative compares with all other alternatives under consideration (i.e., effectiveness is relative among alternatives). Thus, if an innovation is deemed ineffective, it does not imply that the general state of instructional output-student outcomes is ineffective. It simply means that a proposed change is rejected relative to other alternatives which may be further considered or among which choice may be made. Unfortunately, the discussion of ineffective *changes* leads some people to leap to conclusions regarding *general* malaise.

The preceding discussion has seemed warranted because of the record accumulating regarding educational innovations. It is a record of findings of ineffectiveness. The difficulty is not located in the usual place—namely point (3) above and the difficulties of comparative criteria. Most educational innovations can't seem to make it past point (1). Despite changes in instructional outputs, nothing reliably seems to happen to student outcomes!

The Failure of Innovation

The "negative" literature is growing. The Rand study *Federal Programs Supporting Educational Change* (Berman et al., 1975:V, 3-4) states:

> The evaluative research, whose claims to validity are plagued by profound weaknesses in measurement methods, points to rather discouraging general findings: (1) some projects have helped improve markedly some students' skills, behavior, or attitudes, but successful projects are hard to export; (2) few if any projects are consistent—even the most successful ones work well only at a particular time or place, or for some students and not for others.
>
> These results have raised serious questions about the effectiveness of new methods and, in particular, about the usefulness of federal efforts to promote innovation in the schools.

The Rand report suggests four possible explanations for the apparent failure of innovative practices (Berman et al., 1975:I,1):

1. Schools are already having the maximum possible effect; new practices, then, cannot be expected to make a difference.
2. Innovative ideas and technologies tried thus far are inadequate or underdeveloped.
3. Change in student outcomes has occurred, but the measurement instruments are inappropriate or insensitive.
4. Innovative practices have not been properly implemented.

Speculation about these possibilities turns up increasingly in the innovation literature, with the exception of the first. The idea that the schools already are doing what can be done is not appealing to innovators. The other three explanations can be simplified to faulty design, faulty measurement, and faulty implementation. The argument over measurement I leave to others who are more qualified to discuss it. I am inclined to worry about root premises of design as well as the implementation process. Our speculations regarding faulty implementation will await our second model when we examine the interior of the black box. At this point, let's discuss the design problem and briefly speculate about the possibility that the schools might be doing about all that they can.

Roots of the Design Problem

Educational research and innovation may be in difficulty because of four root premises: (1) its traditional paradigm, (2) its mode of research, (3) its treatment of time, and (4) its locus of values. We shall examine each of these possibilities.

Traditional paradigm. The traditional paradigm of education, as presented by McDonald (1975), is illustrated in Figure 4. McDonald comments (p. 5):

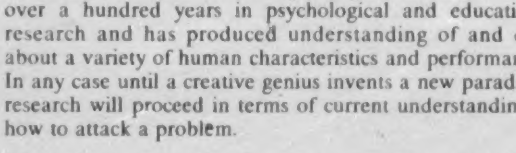
> Any number of specific research designs may be generated within this paradigm. The technicalities of these designs are well understood. Since no one has invented an alternative paradigm, it is impossible to determine whether our modest progress in research on teaching derives from a weak paradigm or from the inadequacies of the research designs that have been used. In defense of the paradigm itself, it can be said that this paradigm has been used for over a hundred years in psychological and educational research and has produced understanding of and data about a variety of human characteristics and performances. In any case until a creative genius invents a new paradigm, research will proceed in terms of current understanding of how to attack a problem.

Well, it may require a creative genius to invent a fully applicable new paradigm, but it does not require a genius to know that this one is wrong. Those who adhere to the fundamental arrogance of this paradigm deserve public campaigns for simplistic teacher accountability! Something much more resembling reality is Cronbach's (1975) concept of Aptitude Treatment Interactions (ATIs).

The one-directional arrows between teacher and student in Figure 4 represent another major case of the one-directional thinking to which Western man is vulnerable (Maruyama, 1968:330):

FIGURE 4

The Paradigm for Studying Teaching Effectiveness


Teacher effectiveness measure
Pupil outcome criterion measure
H
H
L
L

Source: McDonald, 1975

> Western man has traditionally thought of the physical world in terms of cause and effect going in one direction. That is, if A causes B, B cannot cause A. The reason for this assumption is that event order has been confused with logical order; Western man has assumed that because "circular argument" was prohibited in the logic, there cannot be circular causal relationships in the natural or social events.
>
> But not everyone has thought this way. Many tribes in Africa, peoples in pre-Communist China, and some American Indian tribes, especially the Navajos, have seen the universe as a mutual process of various spirits or influences in harmony and occasionally disturbed harmony—in complementary balance rather than in vertical hierarchy. These people have seen the universe in terms of *events* in mutual interaction, rather than in terms of *beings* classified into categories.

Instructional outputs-learning outcomes are *events* in mutual interaction, in my judgment. This is a much more realistic and incisive view than the traditional paradigm based on some *beings* categorized as teachers doing something to other *beings* categorized as students. The punchline is that educational research, innovation, and demonstration projects may be trying to find "a course of reasoning showing that a certain result is a consequence of assumed premises" (the first definition of demonstration) when the *premises regarding causality are false.* The premise of mutual causality is more difficult, but it might lead somewhere.

Mode of research. The mode of research in education is overwhelmingly *reductionist*. In the search for the atomic elements of teaching, behavior has been shredded into behaviors, sub-behaviors, and momentary actions. I believe this reductionist mode of research has carried education as far as it can and may even be serving as an intellectual blindfold.

There are rumblings of discontent within the educational research community. For example, Snow (1974) cites "growing unrest in experimental psychology about what it all adds up to" and describes the knowledge about human learning produced by experimental psychology as "heavily fragmented and task specific." He adds, "Some psychologists, notably those who look to biology rather than to physics for a scientific model, have emphasized anew the importance of ecologically oriented and nonmanipulative research for psychology...." And as a second example, the following beautiful revelation of a researcher's pain and frustration (Berliner, 1975):

> Researchers have spent a good deal of time counting teacher behaviors. We know something about the number of higher and lower cognitive questions asked per unit of time, we have counted the rate of positive verbal praise, the number of criticisms made, the number of probes, the frequency of explaining links, etc. For many of these variables a low correlation with some student outcome measure is found. But in classroom observation one becomes acutely aware of the difference between a higher cognitive question asked after a train of thought is running out, and the same type of question asked after a series of lower cognitive questions has been used to establish a foundation from which to explore higher-order ideas. Teachers sometimes ask inane questions. Teachers sometimes direct questions to what we believe was the wrong child. We have seen positive verbal reinforcement used with a new child in the class, one who was trying to win peer group acceptance, and whose behavior the teacher chose to use as a standard of excellence. We watched silently as the class rejected the intruder, while the teacher's count in the verbal praise category went up and up and up. Teachers have been seen responding to student initiated questions with irrelevant information. Teachers sometimes achieve a high rate of probing student responses to questions, seemingly without regard for the student or the kind of initial response given to a question. Some students are embarrassed by the probing, with other student probes occurring at inappropriate times, and sometimes probes were not used when the situation seemed to cry out for them. Similarly, skillful probing has been observed.... The teacher's probing questions may have been as skillful as Plato's, but only their frequency was recorded.

Perhaps it is time to put the teacher back together again with students in the classroom with all their multi-dimensional interactive events. More important, I believe serious attention needs to be given to *systemic or holistic investigations of the entire set of outcomes* suggested in Table 1 or some alternative version of that table. Professional attention appears to be riveted upon only the first outcome listed—cognitive and affective behavior—and the approach is overwhelmingly devoted to reductionist examination of intermediate events to further the exposition and use of learning theory. Public education is too broad and important a function for it to be the exclusive applied research laboratory of educational psychology. An alternative strategy for research in public education and perhaps eventual innovations is the hypothesis of some set of final outcomes (such as Table 1), the use of a wide range of disciplines in investigation, and a search for systems that might be more effective in terms of those outcomes.

Treatment of time. The time dimension of educational research and innovation is limited to intermediate events. I suspect this is partly a reflection of the reductionist mode and partly a reaction to frustration. For whatever reasons, the focus is on intermediate instructional outputs and intermediate student outcomes—yearly, monthly, weekly, daily, by the lesson. Such a treatment of time limits the horizon, usually to one cottage, and ignores the series of classrooms throughout a student's school career that was illustrated in Figure 3. Intermediate student outcomes are valid ultimately only as they contribute to final outcomes.

Longitudinal analysis of student outcomes would indicate the effect of the entire service system throughout the entire period of public instruction. Are intermediate student outcomes over time additive? Are there inter-cottage leakages? What is the effect of summer vacations? What is the effect upon a mobile child who changes locations and thus systems? How many adults lack computation skill with fractions simply because none of their chain of cottages taught fractions? How many children are switched from self-pacing one year to lockstep the next, with the lockstep cottage assuming certain prerequisites from the previous year which were never reached at the student's own pace? Our statement of outcomes (Table 1) was purposely stated in final terms to emphasize the real societal dimension of goals and evaluation of results. A great fifth grade does not an education make.

Locus of values. Whether stated explicitly or assumed unconsciously, value assumptions are inherent in the design of an innovation. The purpose of proposing change in instruction is to create movement toward valued outcomes. In the case of innovations designed external to local education agencies, the locus of values on which the innovation is posited also is external—in the designer's head. The locus of values that are actually *operative* to change instructional outputs and student outcomes is in the LEA. Unless there is a convergence of values between those designed into the innovation and those operative within the educational unit applying the innovation, the odds are in favor of failure. The evidence reviewed by the Rand group lends empirical support to this conclusion.

Difficulties of transferability and dissemination, so often noted in the literature, may well have this locus of values problem at their root. It may be necessary to start with values operative within an LEA and design innovations responsively, placing the researchers and innovators in a consultative rather than master-minding role.

Are Schools Having a Maximum Effect?

What meaning might we attach to the results of a decade of innovation? Can findings of ineffectiveness be significant? If *changes* seem to indicate little, what does that suggest about the general *state* of the output-outcome relationship?

For policy analysts who have worked in a number of fields, evidence of ineffectual change experiments is fairly common. It leads to the suspicion that another case of a very common phenomenon has been found. The suspicion is that educational effectiveness in terms of student outcomes related to instructional output may have reached the stage of diminishing marginal returns. Perhaps the craft of instruction, if not already having the maximum possible effect, is at least at the stage where further gains are very difficult to achieve.

One rule of thumb indicator of the onset of diminishing marginal returns is when professionals in a given field are devoting great attention to devising very sensitive measurement instruments (physical or statistical). American educational researchers surely are doing just that. By contrast, in developing countries where an innovation might be enough paper and pencils for all the students, available effectiveness measures tend to be illiteracy rate, books in print, and newspaper circulation. Where the effectiveness measures are very refined, the potential for gains is very small. That is not a law, but it is a pretty useful heuristic for policy analysis.

Our conjecture is aimed at efforts to increase student outcomes related to instructional output—not all student outcomes or efforts to increase them. We are suggesting that diminishing marginal returns may have set in regarding tinkering with instructional delivery in the cottages (classrooms)—at least the cottages where research gets conducted. What might this mean?

It does not mean "state of the craft" knowledge would not lead to improvements in perhaps hundreds of thousands of other cottages. Dissemination of common sense findings may remain very important within the guild. It also does not mean that schools (as contrasted with individual cottages) are already having the maximum possible effect. Systemically, schools are much more than specific doses of in-classroom instructional delivery. The earlier discussion of reductionism and longitudinal analysis is applicable here.

What it does mean is that non-instructional possibilities need to be researched. This point is usually difficult for a profession to grasp; perhaps an analogy will help. Doctors want a better recovery rate for heart attack patients. Better and better surgical techniques are develped, intensive care units established, etc. Finally, diminishing marginal returns sets in, and more and more effort, equipment sophistication, and money are rewarded by smaller and smaller increases in numbers of patients saved. It is time to take a systemic view. For example, programs to create public awareness of symptoms and improve ambulance service can diminish the damage that occurs by the time the patient arrives *at* the hospital and thereby save far more lives than would occur by further sophistication *in* the hospital. Systemically, where does education need to look for "outside" help? Should the next innovations be in parent training, preschool and later? What can be done about the student side of the mutual causality of instructional output-student outcome events?

Even the suspicion that instructional improvement may have reached the stage of diminishing marginal returns is worthy of investigation. And confirmation would not be a disaster. Instead, it could serve to direct research and innovative effort away from the well-trod field of instructional change and into little explored areas of potential major gains. If the target really is improvement in student outcomes, then educational research may have to explore unfamiliar territory to be successful at this point in the evolution of American public education.

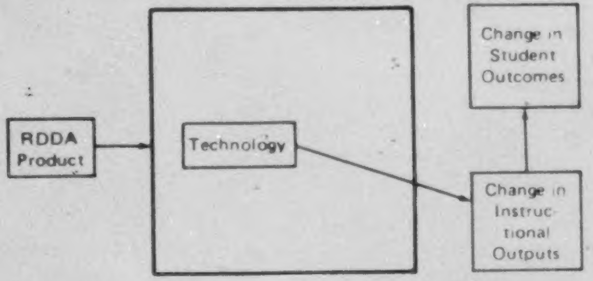
Shifting the Inquiry to Implementation

We have examined briefly possible roots of the design problem and shared our conjecture regarding the possible state that instructional improvement efforts have reached. Our attention now turns to implementation. It is time to examine what happens inside the black box of our first model.

We will begin with an education innovation "product"—machine, material, or method. The product is a result of RDDA, the linear process of moving from research to development, dissemination, and adoption. What happens when this bit of educational technology enters an educational unit? The naive assumption about what happens is shown in Figure 5.

FIGURE 5

Model II
Beginning To Lift the Lid of the Black Box


RDDA Product
Technology
Change in Student Outcomes
Change in Instructional Outputs

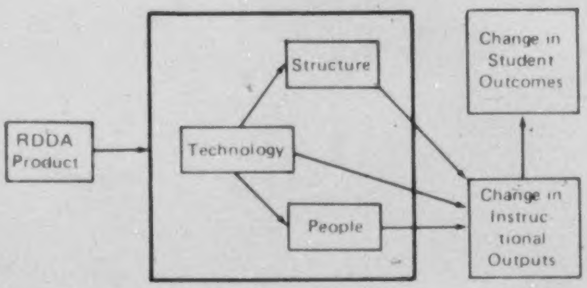
What actually happens to new technology inside the black box? It meets a teacher who proceeds to adapt the product. If it "succeeds," it may well be the "success" of a quite revised product. It also meets a school structure which may facilitate or hinder implementation of the product. And as we've discussed, instructional outputs may change, but we have difficulty finding the concomitant change in student outcomes.

A second view inside the box looks like Figure 6. Figure 6 simply pictures the fact that implementation of a change in technology is *conditional* upon the structural situation and the people involved. If more than single or limited instance success is sought, great attention must be paid these conditional factors. The RDDA model is a fairy tale—after adoption they lived happily ever after.

Innovation efforts, of course, are not limited to use of products. There are those trying to innovate the people through pre- and in-service training of teachers. That route also is conditional—conditional upon the structural situation and upon the machines, materials, methodologies to be used. Others favor innovation in the structure through staff differentiation, class/school reorganization, changes in leadership style, etc. Those efforts are conditional upon the people and the technology.

FIGURE 6

Model IIA
A Second View Inside the Black Box


RDDA Product
Structure
Technology
People
Change in Student Outcomes
Change in Instructional Outputs

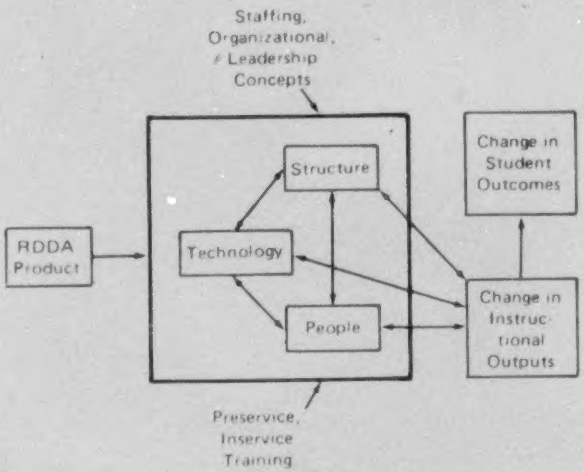
A third view inside the black box is provided in Figure 7. Some will begin to recognize this figure as the result of my taking a good bit of liberty with a model of Leavitt's (1964:317-402) that synthesizes most of the basic streams of development in management thought during this century.

While this is a within the black box view, we have permitted the external innovative drives of the larger professional education community to impact upon the educational unit. Other externalities of a non-professional nature are deferred until our third model.

Figure 7 still uses the traditional one-directional paradigm, so let's add some student interaction in Figure 8 to complete this model. It is essential to recognize that Figure 8 represents a problem-solving model. It assumes that innovation is undertaken to

FIGURE 7

Model IIB
A Third View Inside the Black Box

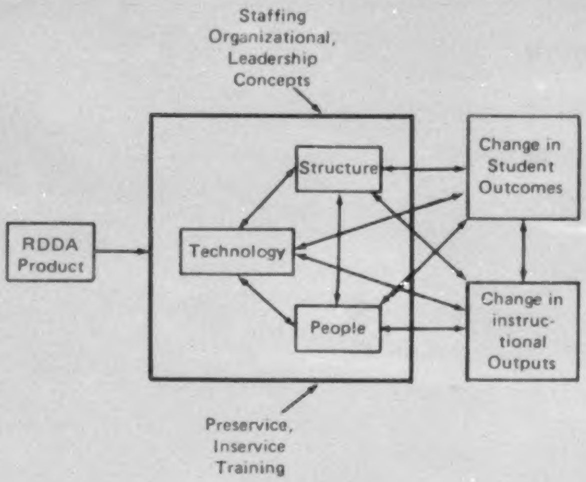

Staffing, Organizational, Leadership Concepts
RDDA Product
Structure
Technology
People
Change in Student Outcomes
Change in Instructional Outputs
Preservice, Inservice Training

change instructional output-student outcome events in order to increase the student gain of "goods." If innovation has other purposes (usually labeled opportunistic), such as solving local budget difficulties, then the problem-solving target is different and this model may have little applicability unless modified accordingly.

The main point of our investigation of this second model is the recognition of the great *variability* that can occur within a public educational unit and the *conditional interdependence* of people, structure, and technology. These factors suggest a high rate of failure for externally developed and packaged proposals for change.

FIGURE 8

Model IIC
Tentative Model
Within the Black Box of an Educational Unit


Staffing
Organizational,
Leadership
Concepts
RDDA
Product
Structure
Technology
People
Change in
Student
Outcomes
Change in
instruc-
tional
Outputs
Preservice,
Inservice
Training

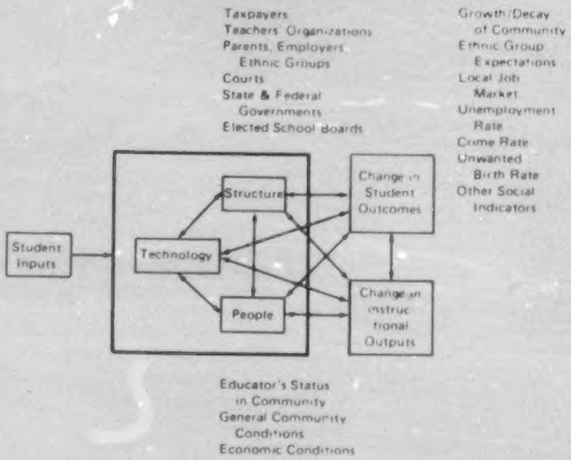
Broadening the Inquiry

Perhaps we have enjoyed the simple conditions of our "in a vacuum" model long enough. It is time to let in all the noise and dirt of the social environment in which public education operates. A first tentative stab at the black box in a context is offered in Figure 9. We shall not belabor this model. Its rough condition hardly suggests that it should be pursued in detail. However, certain observations can be made.

1. All arrows are two-way, indicative of interactive relationships. Arrows of interaction within the local community (outside the education unit) have not been drawn but are considered to be present.
2. The variables interacting with "structure" are sufficient to suggest that the focus of attention within the educational unit may be deflected from problem solving in relation to intermediate instructional output-intermediate student outcome events to solving other problems.

FIGURE 9

Model III
The Black Box in a Context


Taxpayers
Teachers' Organizations
Parents, Employers,
Ethnic Groups
Courts
State & Federal
Governments
Elected School Boards
Growth/Decay
of Community
Ethnic Group
Expectations
Local Job
Market
Unemployment
Rate
Crime Rate
Unwanted
Birth Rate
Other Social
Indicators
Student
Inputs
Structure
Technology
People
Change in
Student
Outcomes
Change in
instruc
tional
Outputs
Educator's Status
in Community
General Community
Conditions
Economic Conditions

3. As a result of a range of conditioning forces, student input is variable. Except under the most extraordinary assumptions, variability of student inputs will result in variability of student outcomes. It is not logical for a distribution of data points all to be above average or any other measure of central tendency. All of the poor cannot have incomes greater than the median income; all non-achievers cannot have better than average student outcomes. The assumption that the process of education should produce absolute equality of student outcomes is similarly extraordinary. It also requires contemplation of a society where, to use an old phrase of Galbraith's, "the bland lead the bland."
4. Student outcomes are not an exclusive consequence of the educational unit. They are a shared product and responsibility of all the elements present in the environment or social context. Both the behavioral and time outcomes shown in Table 1 are consistent with this view. It is unreasonable to assume that a school is a cocoon—an isolated, insulated stage of development—or that the educational cocoon is solely responsible for producing a butterfly.
5. The long timeline of our first model has become longer in the third model. While education is a *now* event, community-school interactions involve memories of school *then*. There are generational differences and difficulties. And while those involved are interested in intermediate student outcomes, education is intended for *tomorrow*, for adult usefulness subsequent to the years of

> schooling. Student outcomes as outlined in Table I are subject to tremendous uncertainty because of the far future periods in which they are utilized.

The main thrust of this third model is to emphasize the large number of variables interacting with the black box and their constraining influence upon the educational unit.

Innovation: Hopes, Chances, and Other "Goods"

Ernest R. House (1975:1) has described the innovator's plight succinctly:

> Transferability. How every heart vibrates to that iron string. Each innovator burns in anticipation of the innovation that will sweep the countryside and attract national attention. Schoolmen clamoring for materials. Teachers grateful for the help. Smiling children working enthusiastically in the classroom. Delicious dreams of preeminence are built on such visions.
>
> But the dreams realized are few or none, the clamor subdued, the teachers somewhat surly.

Our models suggest that the chances of success are slim, indeed. The first model stressed the critical importance of instructional output-student outcome relationships and a record of ineffectual efforts to generate reliable positive changes in student outcomes traceable to innovations. The good we sought we didn't find. The second model found the problem of implementation to be especially difficult because innovation directed at technology, structure, or people is conditional upon the other two and great variability can exist in these elements within and among educational units. The third model found additional variability introduced by the societal environment which tends to constrain what the educational unit can do. Throughout, we have described an industry of thousands of independent cottage-based conglomerates with monopoly characteristics dependent upon guild members pursuing their craft. The opportunities for widespread simultaneous change are not auspicious.

Customary models of research and development, borrowed from high technology, centralized decision-making environments, just don't fit. The educational situation is loaded with variability, interaction, long timelines, and value disagreements. Even if R&D results are achieved, there is no point of decision and structure for implementation—except independent actions by thousands of units permeated by incentives for stability rather than change. It takes great courage and conviction, or naivete, to work in such a vineyard.

Let us side with courage and conviction, professionals dedicated with sufficient fervor to bettering public education to willingly bite such a bitter bullet! Then the situation argues for patience with a lengthy, tortuous process of evolutionary change which *might* pay off in the long run. Such a view also has a short run benefit—financial support for a second cottage industry, the educational innovation community. That is not the kind of "good" we sought initially, but it is good for those so supported.

There are other kinds of "good" around. For example, the Rand study of four major categories of innovation projects provides this most interesting comment regarding bilingual education projects (Berman et al., 1975:V, 17):

> The Rand study indicates that among the four programs, bilingual education projects (Title VII) are the hardest to implement and are the least successful in meeting their goals. Nevertheless, as of 1975, Title VII is the only one of the programs that Congress is willing to support with more and more money each year. Title VII *projects* on the average may not be very effective by the standards of efficiency of innovation, but the *program* has been most effective in legitimizing Spanish-speaking people's demands that the schools pay more attention to their children's needs. In a sense, efficiency has nothing to do with it. There might be a cheaper and more effective way to meet the needs at which bilingual programs aim. But, the political test is potency—the ability of the claimants to win large-scale support from Congress, and thereby the political respect, however reluctant, of school districts that formerly could ignore their demands. Forcing the districts to teach in Spanish is a test of that potency, and thereby contributes to the transcending aim—increasing the social self-respect and political power of Mexican-Americans and Puerto Ricans, children and adults alike.

That also is not the kind of "good" we started looking for—namely, gains in student outcomes—but it is no less a good to those seeking identity, attention, and political prowess!

These two examples indicate that we may have to abandon tunnel vision focused on student outcomes in order to find the "goods" of innovation. Why do educational innovation? Because it pays in diverse ways. By now we have entered the heart of the thicket—the politics of educational innovation. Here is where we are most apt to find demonstration projects.

The basic tactic by the federal government in a number of social programs since the mid-sixties has been to provide financial support for "random innovation." Innumerable organizations have attempted to innovate. The results in other programs as well as education have been limited local success, doubts about transferability, and weak dissemination. Professionals self-helping their own profession have not helped much.

Another approach also has developed: the major experiment or demonstration directly sponsored by the federal government. This approach involves some significant differences from regular random innovations which can be illuminated by returning to our Mercedes-Benz analogy. Characteristically, some theory or collection of theories has developed about a major social problem (the course of reasoning stage).

The major experiment or demonstration is organized (the test track stage) as a test prior to nationwide application of a theory or theories. At this point, our analogy does not hold because the test track is not and cannot be a separable stage. The testing process is *simultaneously* a public demonstration of salient features (the demonstrator stage). Furthermore, it is not a test or demonstration exclusively for professionals; the potential customers among the public can follow what happens through the press. Dissemination, a weakness of random professional innovation, is guaranteed for demonstration. The potential customers can identify the demonstration with their interests—even if the outcomes are perverse from the point of view of involved government bureaucrats. The "science" of demonstration projects can easily become overwhelmed by the emotional manifestations of interest groups. Whereas random innovation can occur more quietly among professionals prior to political decision, demonstration projects are immediately political. Demonstration projects thus tend to be stimulated by ideological or philosophical theories of social remedy. They are political throughout.

How might we compose a scenario for demonstration projects in education? Assume a new federal agency has been chartered to attack the problems of public education. Assume furthermore the usual internal fluidity (chaos?) typical of new federal agencies and the urgency of doing something dramatic in order to survive. What ideological or philosophical theories that could be used as a basis for demonstration projects might our models suggest? Keep the public customer rather than the professional viewpoint in mind. Three possibilities are apparent:

1. Nationwide community memories of school *then*. Regardless of what schools were or did, the perceptions in (some) adult memories reflected in current complaints emphasize cognitive outcomes. Major curricular demonstrations in mathematics, reading, science, etc., are a natural to win the battle of the three R's.
2. The pre-industrial character of education. The military used to run their own arsenals and gun factories but now they have the military-industrial complex. Why not an educational-industrial complex? Demonstrations of contracting out to industry part or all of the task also are a natural.
3. The monopoly character of local education agencies. When there is discontent with a monopoly, the appeal goes out for the market mechanism—competition, demand and supply. Why not let parent-customers buy education of their choice? Demonstrations of a psuedo-market also are a natural.

If an agenda of curricular, contractual, and voucher demonstrations sounds familiar, so be it!

The agency in our scenario may feel compelled to undertake demonstrations due to instincts of survival. To survive, the political content of such undertakings should be understood. When instant politics is involved, the agency needs to recognize the potential limits on its control of demonstrations—on their conduct, evaluation, or subsequent adaptation and implementation (e.g., Rivlin, 1971). Further, the ideological or philosphical genesis of demonstrations potentially gives the agency an image of opposition to values held deeply by the profession affected. That is, the agency runs risks of ire, opposition, and outright attack by the guild. Of course, a demonstration project may lead to an authentic breakthrough. That is the aim—namely, the creation of discontinuity, the initiation of major social change.

Why do demonstration projects? Because the risks to be run are preferable to any available alternative strategy for the agency.

Summary

It is time to summarize this inquiry. What, perhaps, have we learned?

1. Doing demonstrations and innovations is "good"—so long as we are willing to pursue "goods" other than gains in student cognitive and affective behavioral outcomes. Demonstrations and innovations may be a political necessity, with benefits for agencies, interest groups, researchers, congressmen, school boards, superintendents, contractors, publishers, and other beneficiaries not yet discovered. Few public actions keep happening without some good occurring for somebody.
2. The situation for public education is unsatisfactory—for disappointed researchers, surly teachers, and confused students—because the effort has not been clearly productive in terms of effective gains by students traceable to innovations.
3. There are immense difficulties which make the findings of ineffectiveness understandable. These difficulties are outlined by our three models.
4. A conventional research and development model has been tried by education with little positive result. The search for an alternative model is underway.
5. Rather than model patching, we have suggested that a viable new beginning requires the alteration of fundamental premises and methods. Specifically: (a) replacement of the traditional one-directional paradigm with an approach based upon the premise of mutual causation; (b) deemphasis of the traditional reductionist mode of educational research and emphasis upon systemic or holistic investigation; (c) recognition of intermediate outcomes as having full analytic useful-

ness only within the entire chain of outcomes, with consequent attention devoted to longitudinal analysis; (d) identification of LEA's and individual teachers as the locus of operative values, leading to the necessity for a more response-oriented approach by researchers and innovators; (e) investigation of the conjecture that the stage of diminishing marginal returns may have been reached regarding the effects of instructional improvements, potentially leading to quite unfamiliar areas for more productive research and innovation; and (f) realistic appraisal of the difficulties of internal implementation and community-school interaction as critical parts of a successful innovation process.

As an outsider first exposed to the problems and literature of educational research while preparing this paper, I have been impressed with the intelligence, conscientiousness, and courage displayed by those who pursue the systematic improvement of public education. I also have been struck by the almost casual way in which the educational innovation community recovers from failure, hastens to another experiment, and flails away. Perhaps it is time to experiment less and think more.

Notes

1. I am indebted to John Keller for this shorthand method of initiating analytic inquiry.

References

Berliner, D.C. "Impediments to the Study of Teacher Effectiveness." Paper presented for the conference on Research on Teacher Effects, November, 1975.

Berman, P., et al. *Federal Programs Supporting Educational Change*. Santa Monica: Rand Corporation, 1975.

Cronbach, L.J. "Beyond the Two Disciplines of Scientific Psychology." *American Psychologist* 30(2):116-127, 1975.

Dawson, J.E. *Measuring Output as a Basis for Public Budgeting*. Unpublished dissertation, Syracuse University, 1971.

House, E.R. (Untitled.) Paper presented at the National Conference on Innovation and Change, Detroit, September, 1975.

Kogan, L.S., and A.W. Shyne. "Tender-Minded and Tough-Minded Approaches in Evaluation Research." *Welfare in Review* 4(2):12-16, 1966.

Leavitt, H.J. *Managerial Psychology* (rev. ed.) Chicago: University of Chicago Press, 1964.

Maruyama, M. "Cybernetics." In F.J. Lyden and G. Miller (eds.). *Planning Programming, Budgeting: A Systems Approach to Management*. Chicago: Markham, 1968.

McDonald, F.J. "Research on Teaching and Its Implications for Policy Making: Report on Phase II of the Beginning Teacher Evaluation Study." Princeton (N.J.): Educational Testing Service, 1975.

Snow, R.E. "Representative and Quasi-Representative Designs for Research on Teaching." *Review of Educational Research* 44(3):265-91, 1974.

ETHNOGRAPHIC CASE STUDIES IN FEDERALLY FUNDED MULTIDISCIPLINARY POLICY RESEARCH: SOME DESIGN AND IMPLEMENTATION ISSUES[1]

Robert E. Herriott
Abt Associates Inc.
Cambridge, Massachusetts

One of the current priorities in American education at the federal level is to understand better the process of planned change. There seems to be a wide consensus that American society has been changing very rapidly in recent years and that, to be effective as a major social institution, education must discover better ways to respond to such social change at the local, state, and federal levels. An important part of the federal initiative in this respect has been the creation of a series of multi million dollar policy research projects to concurrently stimulate and study planned change within local educational agencies. One of the most ambitious of these has been the Experimental Schools (ES) program initiated in the U.S. Office of Education in 1970 and transferred to the National Institute of Education in 1972.

This paper presents some of the goals and methods of the ES program and of research being conducted under its jurisdiction by Abt Associates Inc. (AAI), an applied social research firm. Major emphasis is given to that portion of AAI research directed at understanding the process of planned change through the mechanism of ethnographic case studies. However, no attention is given to the merits of ethnographic case studies versus more traditional forms of inquiry in educational research. This paper considers exclusively issues in the design and implementation of a particular qualitative approach within a particular large-scale, multidisciplinary policy research project.

The Experimental Schools Program— General Overview

The Experimental Schools (ES) program is part of an important recent change in the character of federal involvement in education. Prior to the mid-1950s, the initiative for educational innovation had been almost entirely in the hands of state and local officials. Towards the end of the 1950s and particularly in the 1960s, federal initiatives increased dramatically. The typical federal approach during this period used the authority of either the National Defense Education Act of 1958 or the Elementary and Secondary Education Act of 1965 to provide a series of discrete categorical grants to states and localities in the hope of stimulating specific curricular innovation. The initial enthusiasm sustaining this approach lessened substantially during the late 1960s, in part because of the demands being placed on the federal budget by the Viet Nam War, but also because evidence began to accumulate that it had not been effective in producing change which persisted beyond the period of specific federal funding.

The ES program arose from a concern that previous efforts had failed because they involved a "piecemeal" approach. ES was conceived as an applied research program to test the effectiveness of a "holistic" approach in which many aspects of a local educational system were to be required to undergo simultaneous change. An important assumption of ES was that the success of a holistic approach did not depend upon the development of new curricular ideas but rather upon the adoption of available innovations in conjunction with a series of structural changes to facilitate their becoming a lasting part of an educational system. To insure a local commitment to the holistic approach, strong guarantees of substantial federal funding over a five-year period were to be made and combined with a system of active federal monitoring of local efforts.

To satisfy the objective of a holistic approach, any project funded by the program was expected to meet a test of "comprehensiveness" by including within its design the following five "facets": (1) a fresh approach to the nature and substance of the total curriculum within a school (or series of schools) in light of local needs and goals; (2) reorganization and training of staff to better facilitate the achievement of particular project goals; (3) innovative use of time, space, and facilities; (4) active community involvement in developing, operating, and evaluating the proposed project; and (5) creation of an administrative and organization structure which supported the project and took account of local strengths and needs.

Between December 1970 and June 1972, three competitions were held to select local school districts and other educational agencies willing to embark on such a program of comprehensive educational change. Eighteen five-year projects were authorized at an eventual overall budget of approximately $40 million. Concurrent competitions were held to select contractors to "document and evaluate" each project. These evaluation projects are on-going with a total budget likely to reach $15 million.

The research component was intended to be central to the overall demonstration. ES actively sought applied research organizations which showed promise of being able to maximize the research opportunities presented by each demonstration project; the analog to each comprehensive local change project was to be a comprehensive research project. The designers of the ES research component argued that traditional program evaluations had proceeded from limited rather than diverse methodologies and often were dominated by the basic assumptions of psychology. They sought the potential contributions of sociologists, anthropologists, economists, and political scientists (as well as psychologists) in their study of comprehensive change. In addition, ES research was to be developed and implemented in a manner which would overcome what were thought to be five major limitations of traditional educational evaluations. In ES research there was to be:

1. An evaluation start up to match project start up—rather than evaluation brought in late in the project's life.
2. Major fiscal commitment to evaluation on the order of 1:2 to program—rather than low level funding resulting in limited types of studies.
3. Major on-site presence for the duration of the demonstration—rather than fly-in, fly-out data collection.
4. Documentation of the local project as a major component of the evaluation—rather than no documentation of what actually was attempted or actually transpired.
5. A major focus on research into the basic nature of holistic change with the purpose of informing knowledge and not simply reporting successes and failures—rather than evaluation commissioned solely as a result of agency or legislative regulation and not from a desire to increase substantive knowledge.

Project Rural

One of the ES competitions to select participating school districts was directed to districts with fewer than 2,500 pupils. From among 320 districts submitting "letters of interest," twelve were selected in June 1972 for the "small schools project." Six of these school districts were awarded one-year grants to plan a five-year project of comprehensive educational change, with a "firm understanding" that they subsequently would be funded an additional four years. The other six districts also received one-year planning grants, but with the clear understanding that long-term funding would be conditional upon the results of their planning process. (This distinction between the two groups was made not for reasons of research design but because of budgetary uncertainty.) Only four of the six districts in the second group eventually received long-term funding. Abt Associates Inc. has had responsibility for studying the first group of six districts since July 1972 and for studying the second group of four districts since July 1973. Within AAI, this research is known as the "Longitudinal Study of Educational Change in Rural America" or, more briefly, "Project Rural."

The AAI research effort has been broadly conceived to address four fundamental questions: (1) What are the sociocultural, political, economic, and historical phenomena of these ten small, rural school districts and their ES projects? (2) What has been the impact of the ES program on pupils, schools, and communities within these school districts? (3) What changes persist beyond the period of federal funding? (4) What knowledge gained through the Small Schools Project is of use to educational policy makers, practitioners, and researchers?

Five separate but coordinated research studies have been designed to contribute answers to these questions. Two of the five are responsive to the *documentation* objective of the ES program. They are tailored to the unique characteristics of small, rural school districts and are conducted individually within each of the ten districts. These site-specific studies rely heavily on anthropoligical and sociological field work and consist of:

1. A series of *Site History and Context Studies* to document how each of the ten communities and school systems developed from their founding to the advent of their participation in the ES program. The report of these studies has been issued (Fitzsimmons, Wolff, and Freedman, 1975).
2. A series of *Ethnographic Case Studies* to document how each of the ten communities and school systems developed an ES project, the problems encountered and solutions reached. The report of these studies will be issued in 1978.

The three other studies address the *evaluation* objectives of the research. They examine all ten research sites in a relatively uniform manner in order to obtain knowledge about those elements of the process of planned educational change which can be generalized to other educational settings. These cross-site studies which generally rely on uniform pre and post administration of standardized pencil and paper instruments, consist of:

3. A *Community Change Study* to evaluate how a rural community and its people, culture, and institutions influence the school system (and its pupils) in the presence of the ES program and how, in turn, the school system (and its pupils) influence the community in the presence of the ES program.
4. An *Organizational Change Study* to evaluate the characteristics of schools and school districts which act as either facilitators or obstacles to the educational change process and the impact of the ES program on the organization of schooling.

5. A *Pupil Change Study* to evaluate the degree to which pupils have been influenced by the initiatives of the ES program and the sources of influence.

Although all five studies are integral to the overall design of Project Rural, the major emphasis of this paper is on the ethnographic case studies.

Design of the Ethnographic Case Studies

The term "case study" has very different meanings within different social science disciplines. In the broadest sense, it means the study of a single case. In psychology, that case is generally a single individual; in sociology, a single organization or community; in anthropology, a single community or social group. The term "ethnography" is used primarily within cultural antrhropology and refers to the researcher's " 'picture' of the way of life of some particular group of people" (Wolcott, 1975). The use of the term "ethnographic case studies" by Project Rural does not imply that the primary thrust of all the case studies is anthropological (several are sociological), but rather that these case studies exhibit many of the distinguishing design characteristics of ethnographies. The most important of these design characteristics are discussed below.

The field worker as the case study director. The Request for Proposals from research contractors called for assignment of a professional staff member full-time to each school district and a small staff of professionals in the "home office" of the organization. Very early in Project Rural planning, it was decided to have the staff member at each site be a professional field worker with primary responsibility for implementing a site-specific case study. Although the content of particular case studies would need to be sensitive to the interactions among community, schools, and ES project, the major mechanism for achieving that sensitivity would be the field worker. It was assumed that those away from the site would be at a great disadvantage in comparison to the resident field worker in making judgments about the relative importance of the variety of phenomena potentially relevant to the fate of an ES project.

The field worker as the data collection "instrument." While the three cross-site studies would use standardized and generally quantified approaches to data collection, the design of the case studies called for the field worker to serve as the major data collection instrument and to document the process of planned educational change through observation, interviews, letters, memoranda, etc.

Each case study site specific. A major and early decision was to deemphasize the role of field workers as gatherers of cross-site data and to capitalize upon their full-time location on site to collect primarily those data which seemed to be most critical to the particular site. A priori comparisons were to be made within the three cross-site studies, and although the field workers were to assist in the collecting data for these comparisons, this duty was not to be confused with their case study responsibilities.

Long-term field work. The Request for Proposals from research contractors called for an on-site presence of approximately four and one-half years, a rather unusually long period for traditional field work. It was decided to design each case study on the assumption that a single field worker would be at each site for the entire period.

Unobtrusive research. The objective of the case studies was to understand the phenomenon of educational change, not to influence it. There is much debate within the social sciences about the desirability and possibility of a researcher's separating himself from the phenomenon under study. Project Rural took a strong stand that the local school districts and their advisors were responsible for producing change and that Abt Associates and its advisors were responsible for understanding change. It was understood that one could not create within these small, rural communities anything comparable to the one-way glass of the psychological laboratory, but the intent was to be unobtrusive to the maximum possible degree.

Holistic orientation. A major assumption of Project Rural was that planned educational change must be viewed in the larger sociocultural context within which formal schooling exists. Such a view draws heavily upon what anthropologists refer to as "social organization" and "world view" but goes beyond the purely social and cultural to consider political and economic phenomena. In the case studies, education was to be viewed as merely one aspect of a sociocultural process which also includes economic pursuits, political processes, institutions such as the family, and voluntary associations such as churches and social clubs.

Induction as a way of knowing. The case studies were designed to emphasize an inductive rather than deductive approach to knowledge. One advantage of placing field workers on site for long periods of time is the opportunity to develop, test, and reformulate theories that are well grounded in the realities of the phenomenon under study. It was the intention of Project Rural to do this through the mechanism of the case studies. Thus, whereas the cross-site studies were to be primarily concerned with testing deductively derived a priori notions about the process of educational change, the case studies were to develop insights from the field experience itself.

Schooling as an alien culture. The design of the case studies assumed that new insights about education and educational change can best be achieved when the field worker brings to the field as few preconceived notions about the structure and function of schooling as possible. Such an assumption suggests that the ideal field worker in Project Rural would be someone educated in a different culture who could view American communities and their schools as a form of "alien culture" whose mysteries could best be unravelled through intensive observation. Although Project Rural never attempted to implement this extreme view of the alien, a major commitment was made to recruiting field workers trained primarily in anthropology and sociology rather than in the study or practice of education.

Previous field work experience. It also was assumed that field workers should have extensive formal training in the social sciences and intensive field work experience. Field work was not seen as something which could be successfully learned on the job within Project Rural or through a period of intensive training between recruitment and placement on site.

Field Worker Role Definition and Recruitment

These design characteristics of the ethnographic case studies were not all explicated in the Request for Proposals nor did they emerge whole cloth in the early months of Project Rural. They were developed through an iterative process of formal role definition and recruitment which was critical to the successful staffing of the ten field worker positions.

The first step was preparation of a formal role definition statement relating the existing sociological and anthropological literature on long-term intensive field work to the Small Schools Project. (See Estes and Herriott, 1973, for the results of this role definition process.) The initial draft of the paper was reviewed by several sociologists and anthropologists experienced with long-term intensive field work and revised in light of their feedback. Concurrently, the project director visited each site to assess local officials' understanding of Abt Associates' research responsibilities and to clarify misunderstandings when they were apparent. In addition, various local documents (maps, newspapers, telephone books, teacher directories, high school yearbooks, etc.) were collected as background information for field worker candidates. Then the senior author of the field worker role definition paper began implementing the role on a pilot basis at one of the rural sites leading to further revision of the paper.

A major recruiting effort was undertaken next through notices in the newsletters of the American Anthropological Association and American Sociological Association and contacts with university departments of sociology and anthropology, field worker training programs, and experienced senior field workers. Curriculum vitae were screened with the nine design characteristics in mind, and each leading candidate was asked to critique in writing the emerging field worker role definition paper. This exercise helped to indicate each candidate's ability to organize ideas on paper and served to immerse the candidates in the peculiarities of this role as compared with the more traditional field worker role, giving them an opportunity to identify personal or professional stresses associated with the role and to consider a variety of ethical issues in the conduct of field work under government contract.

Each leading candidate and his or her spouse subsequently were invited to Cambridge. A series of intensive two-day round-robin interviews were held to consider each candidate's ability to implement the field worker role and to modify the role definition when necessary to remove tensions. Prior to being made firm offers by Abt Associates, each field worker and spouse met with staff members of the ES program and then was accompanied by the project director on a visit to the research site. Although ES Washington staff and local school superintendents had the authority to challenge the suitability of any field worker, no challenges were made.

This field worker recruitment and placement process produced an eleven-member cohort—nine males and one husband and wife team. Seven had their major professional training in anthropology, three in sociology, and one in educational administration. Two held the doctorate at the time of their employment by Abt Associates, seven had completed their dissertation research and much of the writing, while two had yet to begin dissertation research. Of the nine who had completed their dissertation field research, two had done it in a foreign culture, two in Alaska, and five within other parts of the United States. Only three of the field workers had done their dissertation field work in an educational setting—one in a Bureau of Indian Affairs school, another in an alternative school, and the third in a public secondary school. Embodied within these eleven field workers was the design of the ethnographic case studies of Project Rural.

Since July 1972, Project Rural has accumulated approximately 400 person-months of full-time field experience. All the field workers have lived full time on site at least 24 months, and three, for as long as 42 months. There have been no resignations, although in one case it was agreed to accept a field worker's request to relocate temporarily away from the site in order to reduce tensions which were developing between him and the superintendent. Effective August 31, 1976, all full-time field work on site ended, but the writing of the case studies will continue part time for a minimum of twelve additional months.

Some Stresses and Strains

Ethnography as a form of scientific inquiry originated primarily within cultural anthropology. Its basic methodology and certainly most of its research traditions have been developed through the study of groups of pre-literate societies by researchers going into the field on their own or with modest grants from foundations, museums, or universities. During the past three decades, the method has been applied increasingly to the study of groups within contemporary American society, generally by researchers still working alone and without major financial support.

What is being tested in Project Rural is the adaptability of a traditional ethnographic approach to the study of a series of research sites linked contractually to a federal agency and indirectly to a research organization through its contract with the same agency. Thus, instead of a situation in which an ethnographer goes to a site of his own choosing to study a naturally occurring phenomenon, field workers were placed in preselected communities as employees of an applied social research organization under contract to produce a series of research products. Inherent in such an innovative situation are stresses which need to be better understood by those federal agencies who commission ethnography as a form of applied social research, by those research organizations who organize and manage it, and by their professional employees recruited to do the actual research and writing.

Many of the Project field workers have contributed first person accounts to the professional literature to communicate some of these stresses (Burns, 1975; Clinton, 1975, 1976; Colfer, 1976; Firestone, 1975; Firestone & Wacaster, 1976, and Messerschmidt, 1975). In addition, the Project's case study coordinator has made a formal presentation on organization and management (Fitzsimmons, 1975). Examples drawn from their experiences and from the overall project are offered here as a context in which to make some suggestions about a more effective design for ethnographic case studies in federally funded multidisciplinary policy research.

Documentation vs. evaluation. The contract which supports the research of Project Rural stipulates as its major objective the "documentation and evaluation" of this portion of the ES program. Although the evaluation objective is in no sense an issue within the project, the utility of the term itself is. Educators seem to exhibit an extreme amount of reactive behavior in conjunction with the term "evaluation" (Wolcott, 1975). In an attempt to buffer our field workers from the charge that they had been sent as "spies" to feed back to "Washington bureaucrats" the "inside dope" with which to pressure local educators, we tried to make clear in our pre-placement site visits that Abt Associates took seriously its obligation to study rather than influence the local projects. Project staff emphasized that the responsibility for making evaluative statements resided in Cambridge and would be exercised primarily through the three cross-site studies and other summative reports. The field workers were not to be in direct contact with federal officials nor were they preparing reports for Cambridge which would be passed on to federal officials responsible for monitoring or funding local projects.

Despite a concerted effort to avoid having the field workers tagged with the term "evaluator," it has plagued several of them throughout their period in the field and created numerous problems of rapport in areas relevant to the case studies. Particularly troublesome has been a tendency for representatives of the federal government with responsibility for monitoring or reviewing local projects to confuse the documentation and evaluation responsibilities within Project Rural. This has occurred primarily through unintentional communication to local school personnel that the field workers report directly to federal officials. Such actions–whether intended or unintended–often undermine the ability of a field worker to establish and maintain the rapport which traditionally has been at the heart of the ethnographic method. (See Burns, 1975, and Colfer, forthcoming, for more extensive discussions of this problem.)

Local confidences vs. federal confidences. Because of the competition among federal agencies for scarce funds, there seems to be a necessity for repeated justification of the continued existence of complex long-term research projects and their component parts. This generally produces stresses between the project and its sponsor and often among the various components of the project itself. Within Project Rural, there has been tension over the appropriate resource balance between the ethnographic case studies and the three cross-site studies. The case study directors (i.e., the field workers) have been at a disadvantage in this competition because of their geographical dispersion and the lack of long-term experience with traditional ethnography on the part of both the research organization and research sponsor. Particularly troublesome has been a concern of the field workers that if they shared the intermediate products of their research–their field notes, interview protocols, research diaries, informal working papers–with staff members of AAI and the research sponsor who have no responsibility for local project monitoring and funding, these products would fall into the hands of those who are responsible for funding and monitoring and lead to the impression–if not the actuality–that the field workers were in fact "spies for the Feds."

The field workers have been caught in a double bind. If they willingly share the intermediate products of their research before completion of the contractual relationship between the federal agency and their research sites, they run the risk of inadvertent–or deliberate–premature disclosure with the consequence that they will be shut off from major data sources. On the other hand, if they fail to share their intermediate products, they run the equal risk that the federal agency will conclude that the case studies are unproductive. In this case, their research is likely to be terminated prematurely by the research sponsor rather than the research site.

This double bind has led to frequent and repeated negotiations between the project director, officials of the research origanization, representatives of the federal agency, and the field workers. Each time there has been a change of personnel within the federal agency, the issues have had to be dealt with anew. Throughout the negotiations, Project Rural has attempted to retain the on site viability of its field workers, but on occasion risks have been taken in order to keep the overall project viable. One compromise was an agreement to having an intermediate case study product reviewed by an employee of the federal agency who had no responsibility for the ES program. This action seemed to allay temporarily fears within the federal agency that the case studies were unproductive. However, when the key officials of one school distric learned of this, they declared our field worker "persona non grata" and caused him to leave the field after only two years. (See Messerschmidt, 1975, for a spirited discussion of this problem.)

Field initiated vs. centrally mandated responsibilities. One of the most pervasive tensions has been associated with the field workers' responsibility to carry out obtrusive data collection activities for the cross-site studies (or overall project management) in conflict with the unobtrusive posture of the enthnographic case studies. Although the necessity for field workers to participate in cross-site data collection was made explicit and accepted by all candidates, neither field workers nor Project leadership fully anticipated the differences among sites that made it extremely difficult to respond uniformly (in terms of both substance and timeliness) to requirements for cross-site data. What might be a simple task at one site because of public records which could be reviewed quickly and unobtrusively could turn out to be a major crisis at another. For example, a question about the frequency of unwed teenage mothers was not a problem at most sites but could have seriously jeopardized case study capabilities at another if it had been asked of the suggested informant—a local health official who in this case was simultaneously chairman of the school board and the parent of a teenage unwed mother.

Because the field workers entered the field from four to fourteen months after the local projects began and generally several months after completion of cross-site study design activities, they were under great pressure to collect obtrusive baseline data during the very period when field workers traditionally have been advised to maintain a low profile in order to establish rapport on site. In some instances, such data collection facilitated the case studies, providing access to data which would have been useful even if there had been no cross-site studies, but in many instances it was in clear conflict with the approach of the ethnographic case studies. (See Clinton, 1975, 1976, and Firestone, 1975, for further elaboration.)

Time for data collection vs. time for writing. An ethnographic approach to educational research is highly "labor intensive" in the time required for both data collection and preparing the research report. Although there is no such thing as a typical ethnographic approach, experienced field workers argue in general that competent field work requires at least as much time for data analysis and writing as for field work. The design of Project Rural required data analysis and writing to take place in the field rather than in the museum or university office typical of traditional ethnography. This made the field workers particularly vulnerable to two types of unanticipated tensions.

On the one hand, it was difficult for field workers to resist ad hoc requests from Cambridge to collect cross-site study data even when the natural rhythm of the case study at that site called for a period of intense review and writing in isolation from the daily activity of the community and its schools. On the other hand, continued on-site presence has on occasion tempted field workers to continue their field work into a period of time more naturally suited for data analysis and writing.

In general, case study preparation has progressed most expeditiously when the field worker made a clear transition after about three years from a primary concern for case study data collection to data analysis and writing. In some instances, making this transition has required the field worker to relocate his residence away from the research site, returning only periodically to facilitate cross-site study data collection or to clarify some particular aspect for the emerging report. In a few instances a field worker has made a successful transition to writing while remaining on site, but this has required an unusual degree of discipline (occasionally obstinance) and judicious extrication from a host of interpersonal relationships. (See Firestone and Wacaster, 1976, for more detailed discussion of the problems of intensive long-term field work in Project Rural.)

Some Suggestions

The suggestions which follow have been drawn from the experience cited illustratively in the preceding section. In considering their utility, one should keep in mind that they have been derived from experience in a project with the following distinguishing characteristics:

1. Project Rural has been embedded within a set of

complex contractual relationships linking a federal contractor (Abt Associates Inc.) and ten program contractors (the rural school districts).

2. Project Rural has been a large research project by conventional standards. Its six year budget is likely to approximate $5 million, exclusive of the approximately $8 million being paid to the ten school districts for planning and implementing their change projects.
3. The ten research sites are highly dispersed geographically, located in rural Alaska, Arizona, Kentucky, Michigan, Mississippi, New Hampshire, Oregon, South Dakota, Washington, and Wyoming.
4. Project Rural has been highly multidisciplinary, involving a Pupil Change Study drawing heavily upon statistics, psychology, and social psychology; an Organizational Change Study drawing heavily upon social psychology and sociology; a Community Change Study drawing heavily upon economics, sociology, and public policy; and ten Ethnographic Case Studies drawing heavily upon anthropology and sociology. The general approach of the three cross-site studies has been to apply uniform, standardized and quantitative methodology to all ten sites; that of the case studies has been variable across the ten sites, unstandardized and qualitative. The suggestions which follow are concerned primarily with improving the viability of ethnographic case studies in research having these characteristics. Although they no doubt have relevance to other types of educational research, appropriate caution against overgeneralization should be exercised.

The suggestions are addressed to three audiences—federal agencies that commission large-scale policy research projects, applied social research organizations that organize and manage them and prospective field workers.

Federal research agencies. Federal research agencies desiring to complement the type of knowledge gained from longitudinal designs involving standardized tests, attitude questionnaires, and sample surveys with that available from unstructured observation, key informant interviewing, and the study of site artifacts should consider some major changes in the research design and implementation decisions made by the Experimental Schools program.

1. *Earlier start-up of field work.* Although the ES program made a major advance in having the research and programmatic efforts begin concurrently, there is a need for even-earlier entry into the field. Successful ethnographic field work seems to require that the field worker establish credibility well *before* local citizens, school personnel, and pupils are impacted by federal project monitors and the array of obtrusive data collection instruments associated with cross-site studies. The field workers in Project Rural entered the field from three to fourteen months *after* the projects began their contractual relationships with the federal agency and in general only two months before the first cross-site study data collection. In retrospect, this seems too late on both counts. Given the pervasive federal practice of selecting both research sites and research contractors in the waning hours of each fiscal year, it may be difficult to fund research contractors prior to selecting the research sites. However, efforts can be made to encourage the research contractor to avoid a rush to collect obtrusive baseline data during the first months of on-site field work.
2. *Early clarification of case study audience.* There has been a tendency on the part of the sponsors of Project Rural to feel that an ethnographic case study can be all things to all people—that it can speak simultaneously to the policy maker about what legislation to draft or programs to implement, to the practitioner about how to organize and manage change, to the citizen about how to participate more effectively with professional educators in the change process, and to researchers interested in achieving a better understanding of schools and schooling. Such is not the case.

   Traditional ethnography has been written primarily to advance social science knowledge. The experiences of Project Rural suggest that it can be adapted to serve other audiences as well, but primarily if those audiences and the associated case study goals are specified before the recruitment of field workers. If the sponsoring agency has or is likely to develop a strong preference for a particular applied audience, this preference needs to be made explicit very early so that field workers can be recruited and placed on site with those expectations clearly in mind. Although the field worker is not as greatly constrained in his ability to make "mid-course corrections" as, for example, the social psychologist locked into a longitudinal design involving uniform standardized tests and control groups, there are serious constraints. In ethnographic case studies the field worker is, in fact, "the instrument," and research sponsors should approach mid-course corrections in the types of problems to which these instruments are to be applied with the same degree of caution they would use in considering different standardized instruments for the pre-and posttests of a longitudinal study.
3. *Peer review of intermediate case study documents.* In the absence of extensive experience within a

federal agency in the nurturing of ethnographic case studies, it seems essential that procedures be established to buffer field workers from the inherent conflict between federal and local confidences discussed earlier. Once the sponsor has clarified its preferences regarding the primary audience and supervised the recruitment and placement of qualified field workers well in advance of other forms of data collection, it should select an external panel of experienced field workers to judge the relevance, quality, and utility of the field approaches employed and the documents being prepared intermediate to the final case study report. Ideally, these panelists should be appointed immediately upon the initiation of field work, but in any event, prior to the sponsor's first need to assess case study direction or progress.

In Project Rural, such a panel was established informally by Abt Associates when field workers were recruited and then formalized after the first year of field work. Even though the panel's potential membership was reviewed by the sponsor, its appointment by the research contractor seems over time to have undercut its credibility as a group able to speak in the best interests of the sponsor. At present, Project Rural is faced—three and one-half years after case study field work began—with the necessity of developing a productive relationship with a new case study panel selected exclusively by the sponsor.

4. *Clear distinction between program and research monitoring.* Agencies responsible for monitoring related program and research contracts often have difficulty maintaining appropriate distinctions between the two. Initially, the ES program made a sharp distinction between these two types of monitoring responsibilities with separate project officers. However, as travel funds became scarce and particularly as resignations occurred during hiring freezes, there was a tendency to double up program and research monitoring responsibilities in the same project officers. This greatly raised anxiety among local school personnel and field workers that confidentiality agreements hammered out at great cost in time and dollars no longer existed or at least were being renegotiated. It also created conflicts of loyalty when monitors were unable to change hats effectively when moving from one responsibility to the other. Because of the fragile relationship between case study success and local confidences, the monitoring of research contracts involving ethnographic case studies of the type in Project Rural would appear to be best served by a consistent and pervasive distinction between program and research monitoring.

5. *The sponsoring agency as an important research phenomenon.* The Request for Proposals creating Project Rural spoke only of the local school districts as important research sites requiring on-site presence. No mention was made of the possibility that events in Washington would have important implications for the fate of the various local projects. When the importance of Washington as a research site first became apparent to the research contractor, several proposals were made for increasing the contractor's ability to understand this phenomenon. In each instance, considerable resistance was encountered from the then leadership of the Experimental Schools program. It now seems that the research objectives of the program could have been greatly enhanced if the ES program itself had been declared an eleventh "site" with a full-time field worker present throughout the life of its contracted relationship with the ten school districts. The resulting ethnographic case study could have added much to our understanding of the process of federally stimulated local educational change.

Applied research organizations. Research organizations which contract to organize and manage ethnographic case studies need to be particularly sensitive to a rather unique set of personnel matters associated with this form of research. The skills essential for functioning effectively in field settings are not necessarily the same as those essential for success within a sophisticated research organization.

1. *The control of time.* Particularly troublesome are tensions over the control of time schedules. Within sophisticated research organizations, time is highly structured and generally scheduled in advance. In the field, time is much more elusive, and work schedules must adapt to the fact that the events under study are not under the field worker's control. Such distinctions in the control of time exist between any field site and central office. When there is a series of field sites in dispersed geographical areas, the problems of time control and coordination compound, especially given the natural desire of a central office to receive communications from the field uniformly.

   It would seem essential in such situations that the research organization create a role in the central office to buffer all communications between field workers and the rest of the organization. The incumbent of this role needs to have sufficient power within the organization—and knowledge about on-site conditions—to argue effectively with those less sensitive to the peculiarities of field work.

2. *The importance of field worker selection.* Although careful buffering of field workers is essential, it will mean little if care has not been taken in the recruitment and on-site placement of the field workers. The strategy employed by Project Rural was very expensive and time consuming, yet it seems to have succeeded in avoiding a problem of field worker turnover which could have been detrimental to the objectives of the ethnographic case studies. All the buffering one can muster will be for naught if recruitment and on-site placement has been done carelessly.
3. *Protection of case study resources.* Research organizations which have contracted to carry out traditional ethnographic case studies within multidisciplinary projects need to aggressively seek ways to insulate contractually the case study component from the effects of budget crunches within the sponsoring agency. Each change of leadership within sponsoring agencies seems to bring new priorities which must in some sense be acknowledged by their research contractors. Such changes are particularly troublesome for longitudinal studies under annual funding (as Project Rural has been since January 1974). In such situations, studies dependent upon pre- and post-designs using standardized instruments seem to have a great advantage in the struggle for survival. The apparent rigidity of these designs seems to offer greater protection from redefinition than is true for ethnographic case studies, given the latter's elusive character and the evolutionary nature of their development.

Prospective field workers. A special form of entrepreneurship seems to be essential when traditional ethnography is carried out under complex contractual conditions. In traditional ethnography, entrepreneurship consisted primarily of negotiating access to a particular group of people who were to be the subjects of the research and then avoiding the violation of a variety of taboos. Throughout all of this, the ethnographer was very much on his own. Under complex contractual conditions, survival on site is still essential, but it is complicated greatly by the need for frequent negotiations within a research organization and between it and its sponsor. When sponsor, research organization, and research site are linked contractually some pretty agile footwork often is required.

1. *Careful assessment of research organization and sponsor.* The field worker needs to be confident that the research organization and research sponsor have a realistic sense of the complexities of the field role. The role definition statement in Project Rural went a long way towards engendering that confidence. The fact that the field workers participated in its revision and that it was reviewed and accepted by the research sponsor also seemed to help. All prospective field workers should insist upon a process of written role definition, field worker review, collective revision, and sponsor sign-off.
2. *Continual vigilance.* An a priori written statement, however, can only go part way in ameliorating the potential stresses. Any weakness in the Project Rural field worker role definition seems not to be in the fact that it was unimplementable, but in the fact that it failed to anticipate the fragile nature of even *contractual* commitments between research sponsor and research organization. Each time personnel changed in the sponsoring agency or a review of its effectiveness was mandated by higher administrative levels, new actors would come in contact with the field workers (and people at their research sites)—actors who were not necessarily privy to previous understandings and who often neglected to act in ways consistent with them. Since there seems to be little a research contractor can do to anticipate all new initiatives between a sponsor and its research sites, there seems to be no alternative but for all field workers to be continually vigilant to the inevitability of inadvertent violations of apparent understandings between their research organization and its sponsor.

Notes

1. This paper should not be construed as a report from Abt Associates Inc. to the National Institute of Education. The observations and suggestions contained herein, although influenced by the author's official responsibilities, are offered in his private capacity as a sociologist interested in the organization of educational research.

References

Burns, A. "An Anthropologist at Work: Field Perspectives on Applied Ethnography and an Independent Research Firm." *Council on Anthropology and Educational Quarterly* 6:28-34, November, 1975.

Clinton, C.A. "The Anthropologist as a Hired Hand." *Human Organization* 34:197-xx, 1975.

Clinton, C. A. "On Bargaining with the Devil: Contract Ethnography and Accountability in Fieldwork." *Council* *on Anthropology and Education Quarterly* 8:25-28, May, 1976.

Colfer, C.J.P. "Rights, Responsibilities and Reports: An Ethical Dilemma in Contract Research." In *Ethics and Anthropology: Dilemmas in Field Work*, P. Spradley and M. Rynkiewich (eds.). New York: John Wiley, 1976.

Estes, R.P., Jr., and R.E. Herriott. "The Role of the On-Site Researcher in Project Rural." Unpublished manuscript, Abt Associates Inc., November, 1973.

Firestone, W.A. "Educational Field Research in a 'Contract Shop.' " *The Generator* of Division G. American Edcuational Research Association 3: Spring, 1975.

Firestone, W.A., and C.T. Wacaster. "The Promise and Problems of Long-Term, Continuous Field Work." Paper presented at the Annual Meeting of the American Education Research Association, April, 1976.

Fitzsimmons, S.J. "The Anthropologist in a Strange Land," *Human Organization* 34:183-96, 1975.

Fitzsimmons, S.J., P.C. Wolff, and A. Freedman (eds.). *Rural America: A Social and Educational History of Ten Communities*, Vols. I & II. Cambridge (Mass.): Abt Associates Inc., 1975.

Messerschmidt, D.A. "Problems of Role and Method On Site: A Case Study." Paper presented at the Annual Meeting of the American Anthropological Association, November, 1975.

Wolcott, H.F. "Criteria for an Ethnographic Approach to Research in Schools" *Human Organization* 34:111-27, 1975.

CRITIQUE

Walter J. Symons
Alum Rock Union Elementary School District
San Jose, California

In making some observations about the papers by Herriott and Dawson, I will limit myself to the relationships between the ideas they presented and the operational realities within a public school system. It is exactly those "operational realities" that represent the ritualistic hurdles that are in the way of progress in public education. I would add, however, that the operational realities of research and of the formal and informal power structures within the agencies that fund this research also are significantly adverse to the progress of public education. I don't believe either Herriott or Dawson would disagree with these assumptions.

The study described by Herriott illustrates some of these problems. In my mind at least, there is no question that qualitative research, especially using ethnographic methodology, is what we need in public education. The most effective way to measure *what* is occurring is to get into the classroom and so describe it that it can be analyzed and used for future improvement. One of the problems, however, is that the field worker performing ethnographic research "where the action is" is always going to be viewed as a "spy," as an evaluator. There is a consistent effort on the part of researchers, and this includes Project Rural described by Herriott, to convince the research site that the ethnographer is not, in truth, an evaluator. To me, this is like saying, "The Emperor is fully dressed and looks beautiful in his new clothes." How does one observe and then describe this observation without its being viewed as evaluation? An observation is simply the comparison of some relationships, not necessarily good or bad, but distinctive enough to be recognizeable. In short, I don't believe it is possible to do anything that would insure acceptance of an observer as a nonevaluator, even if the observer were recruited from within the ranks of the research site.

It is important, however, in making it possible for ethnographic observers to work successfully on site that the agencies conducting and funding the research not make value judgments before collecting data. It has been the general rule among the researchers with whom I have worked recently that they demonstrated their biases before collecting and analyzing data and drawing conclusions. If Herriott's statement is true that "educators seem to exhibit an extreme amount of reactive behavior in conjunction with the term 'evaluation,' " this premature display of value judgments may be one of the reasons.

The field workers' problems as described by Herriott would seem to indicate a need to reexamine the field worker's role and the process surrounding his participation in ethnographic educational research. One problem concerns the time at which ethnographers enter a project. Herriott recommends that it be much earlier than was the case in Project Rural—a recommendation confirmed by my own experiences with the Voucher Project in our school district. The ethnographer in our study was assigned at the beginning of the demonstration and was viewed as an integral part of the overall set of relationships necessary for conducting a successful project. But even in this case, many difficulties arose because he had not been a participant during the preceding year when a feasibility study was conducted and much of the planning for the project accomplished.

While there are obvious risks in recruiting ethnographers from within the research site, one advantage is that they can be part of a project from the very beginning. Moreover, Herriott's comment that traditional ethnography has been written primarily to advance social science knowledge supports, in my view, the contention that someone more familiar with the cultural setting, who is well trained in his role, could be a more productive observer than an outsider. It would seem that a far more accurate set of perceptions would be brought to the observation if the ethnographer had a background with which to understand the nuances of politics on the research site, traditional policies and practices, and the participants who were to be observed.

The section in Herriott's paper on local confidences vs. federal confidences hit some sensitive nerves. It seems to me that the greater the national, or even local, interest in a research project and the broader the aura of possible contribution to the research community, the greater the appetite for control is likely to be. The formal and informal power structures within agencies, both local and federal, only aggravate this tendency. Suffice it to say that we at the research site are just as willing to agree to anything to get the money as are the federal agencies willing to agree to anything to have a research site accept a project, especially if it is highly experimental.

While there are no ready answers for these problems, they are sufficiently serious that it behooves any project to define contractual responsibilities (including those of ethnographers) carefully, to provide appropriate inservice role training for the participants, and to gather the participants together to deal with some of their biases so that the project can operate with greater trust on the part of all.

Dawson's paper, "Why Do Demonstration Projects?" intrigues me when he responded, "Why, indeed!" Probably we all have asked ourselves the same question, especially when caught in a demonstration beset by difficulties.

In his discussion of the analyst's "three great questions," Dawson implied a concern for finding the "goods" in demonstrations. We may need to devote much more time to looking for *what* is happening rather than hypothesizing that one thing is "good" or better than another. This is not necessarily easy to do. We viewed the Voucher Project as an opportunity to discover what would happen when a traditional urban educational system introduced some major changes into the existing roles and traditions. Although everyone involved had agreed to this view, the fact was that the research sponsors hoped to demonstrate the "good" things they secretly desired, while opponents were certain the entire experiment would lead to "no good." Institutions seem to have tremendous difficulty collecting a reasonable group of participants who are willing from beginning to end simply to say, "Let's see what happens," and then decide whether they have increased the alternatives from which to choose.

The Dawson paper also causes me to comment that we need to take a serious look at the effect broad societal expectations are having on the public schools. I am not at all convinced that the results of demonstrations should necessarily reflect some growth on the part of the student. Dawson's concept of the black box in a vacuum illustrates the point that public outcomes can be a product of demonstrations. By weighing the effect of the many incidental agencies and individuals that are considered necessary to a school system, we may find far more valuable information for resolving educational problems than we will by limiting ourselves to a narrow focus upon students. The observation in the Dawson paper that "despite changes in instructional outputs nothing reliable seems to happen to student outcomes" supports the idea that we may be spending far too much time measuring the wrong things in educational innovations.

I would call your particular attention to the quotation in the Dawson paper from the Rand report which suggests four possible explanations for the failure of innovative programs in education. If we could sit down with each of the school systems which provided the data that prompted these conclusions and ask, "Why were these conclusions drawn about your schools?" we might get at the heart of the reasons for failure in educational innovation.

Dawson's paper was lucid in its questioning of present and past practices of limiting the measures of educational outcomes in demonstrations. For me, he sung out that the Emperor (meaning demonstrations) was naked, indeed! As he concluded, we may need to "experiment less and think more."

CRITIQUE

Gene E. Hall
Research and Development Center
for Teacher Education
The University of Texas at Austin

In reviewing the Dawson and Herriott papers, I would like first to discuss each individually and then develop a few general points.

In his paper, Dawson systemically developed an insightful analysis of the problems and dilemmas of contemporary research in education. He raises several implicitly held assumptions that need to be questioned more seriously by qualitative and quantitative researchers alike as well as by policy makers.

One of these assumptions concerns the reasons for undertaking innovative projects. Dawson's general conclusion is that we do "innovations" because they have useful political or economic ends. Unfortunately, it seems that all too many projects are done for exactly the reasons Dawson proposes.

Another set of issues arises in connection with the variables that are studied. In discussing his Figure 3, Dawson makes the point that research needs to cover a longer period of time than six months or a year of schooling. As Figure 3 displays and he points out, this longer range perspective makes for tremendous variation. A very large number of variables would need to be taken into account before generalizations could be made, or the sample would need to be enormously large before conventional statistical tests could be performed.

In recent years, researchers have focused on using many different variables as covariants. Most of these "aptitude treatment interaction designs," however, are set up with only one criterion variable. One of the implications of Dawson's Figure 3 is that we need to be looking at multiple criterion variables just as we have been looking at multiple predictor variables. Each outcome individually can be related to a certain combination of the covariants, but in addition, the multiple outcomes *in various combinations* will represent additional variables that need to be studied. All of this raises questions: Do sufficient statistical tools exist for these analyses? Does it make sense to keep emphasizing single criterion variables? What are the funding implications of extended longitudinal studies?

Equally important, we need to stop clinging to those age-old variables in educational research that have not demonstrated any new relationships. Perhaps through qualitative research methodologies and new theoretical constructs we can identify new variables which, when studied with our existing precise methodologies, will yield new understandings.

Elsewhere, Dawson suggests that "most innovations can't make it past point 1," point 1 being "whether a change in a controllable element (an innovation in the production of instructional outputs) induces a change in outcomes." My own research indicates that most research studies actually do not make it *to* point 1. For example, in one evaluation study we documented, 49 per cent of the teachers in the so-called "control group" were using the innovation, while only 84 per cent of the teachers in the "treatment group" were using it. Depending on whether the data analyses were done on the control group versus the treatment group or on users versus nonusers, the outcomes were completely different. Faith in the sampling design will not suffice; we need to make a valid check of whether or not implementation has occurred.

Finally, I would suggest that the Dawson paper needs an additional figure. We know from research that all the vectors in his Figure 9 are not the same size. Some carry a great deal more weight and account for more of the variance than do others. The teacher in the classroom, for example, accounts for much more of the variance with regard to the degree and effects of implementation than do many of the other factors and conditions outlined. An additional figure would show vectors with differnet lengths to reflect more closely our knowledge of how these variables interact.

Turning to Herriott's paper, from which I gained a number of helpful ideas, there are several points I want to make. One of the most important and exciting ideas reported is the procedure of assigning a full-time field worker to each site. This is a long way from the one-shot posttest evaluation design, and it offers the opportunity to gather extensive information within a field setting. The strategies Herriott describes for selecting and training these field workers sounds most effective—especially the powerful technique of asking prospective field workers to critique a concept paper about their role.

I am, however, uncomfortable with the attempt to make the field workers unobtrusive by not allowing them to be actors in the system they were studying. As Herriott documents, the end result was that the field workers were anything but unobtrusive. It seems to me they would have been less obtrusive and less threatening if they had been participant observers, playing a complementary role within the client system.

It also would appear that the failure to provide feedback to the subjects only increased the difficulties between field workers and clients. Given the length and size of the study, there surely must have been some information that could have been fed back to the subjects. Almost any data would at least have given the client system a feeling for the kinds of data being collected and helped to establish more credibility for the field workers and the study itself.

Herriott's suggestion that "aliens" to the school culture be used as field workers is interesting. I suspect, though, that such "aliens" would need to be recruited from outside the United States, from countries in which the schools differ from American schools. Americans outside the field of education are not apt to be any less biased or knowledgeable about American schools than are educationists. It would be informative to know whether or not the "aliens" did, in fact, differ from trained educationists in the kinds of data they collected.

Herriott placed considerable emphasis on the problems of contract research. I would observe that all of us who have been involved in federally funded research have been through the experience of continually changing priorities and project monitors, crisis requests for products, do-or-die site reviews, and continuing uncertainty about funding. This is one of the realities of federal contract research and not at all unique to qualitative research methodologies.

I will conclude with two points relevant to both papers. First, it would seem that if Dawson is right in terms of the complexities of output and outcome variable identification and analysis when studying demonstrations, the multivariable qualitative design of Herriott is right on target. On the other hand, if Dawson is correct that significant changes in output and outcome variables are not likely, that we've pushed schooling in its conventional form about as far as it can go in making significant changes, then the whole thrust of the Experimental Schools Program, of a big dollar attempt to make small changes in a few rural schools, is probably not a sound investment.

Second, Herriott's whole project flies in defiance of the political necessity for quick glory. Dawson's paper and Herriott's work suggest that we need to pursue more of these long-term studies. Yet, the press for political and economic gain seems all too real, and annual funding and the shifts in policy occassioned by elections all add up to there being less than the critical amount of stability required for long-term research. Both authors would agree, I suspect, that we have some fundamental problems to resolve if the question, "Why do demonstration projects?" is to be answered positively.

HOW TO IDENTIFY EFFECTIVE TEACHING

The study of teaching to date has relied on rather narrow definitions of what is effective teaching. Such definitions most usually are tied to student outcome measures such as achievement tests, while focusing on individual, isolated teaching acts rather than on teaching and learning in its total context. How can we identify effective teaching, recognizing the total ecology of the teaching-learning environment?

HOW TO IDENTIFY EFFECTIVE TEACHING

William W. Cooley, Gaea Leinhardt,
and Janet McGrail
Learning Research and Development Center

In this paper, we will describe how quantitative research methods can contribute to the identification of dimensions of effective teaching. By identifying such dimensions and demonstrating that they are reasonably stable, eventually it will be possible to achieve more valid assessments of individual teacher performance. Our focus is on the research task of establishing general features that characterize effective and ineffective teaching rather than on the clinical task of making decisions about individual teachers for such purposes as certification. We feel that a great deal of work remains to be done on the problem of identifying effective teaching before procedures and instruments for identifying effective teachers can be significantly improved.

We also want to emphasize that although we have been asked to take a quantitative point of view, we do not intend to argue the superiority of quantitative methods over other research methods. Doing so only results in the kind of name calling that Meehl (1954:4) summarized over twenty years ago in his analysis of the debate surrounding the merits of statistical versus clinical methods:

> It is customary to apply honorific adjectives to the method preferred, and to refer pejoratively to the other method. For instance, the statistical method is often called operational, communicable, verifiable, public, objective, reliable, behavioral, testable, rigorous, scientific, precise, careful, trustworthy, experimental, quantitative, down-to-earth, hardheaded, empirical, mathematical, and sound. Those who dislike the method consider it mechanical, atomistic, additive, cut and dried, artificial, unreal, arbitrary, incomplete, dead, pedantic, fractionated, trivial, forced, static, superficial, rigid, sterile, academic, oversimplified, pseudoscientific, and blind. The clinical method, on the other hand, is labeled by its proponents as dynamic, global, meaningful, holistic, subtle, sympathetic, configural, patterned, organized, rich, deep, genuine, sensitive, sophisticated, real, living, concrete, natural, true to life, and understanding. The critics of the clinical method are likely to view it as mystical, transcendent, metaphysical, supermundane, vague, hazy, subjective, unscientific, unreliable, crude, private, unverifiable, qualitative, primitive, prescientific, sloppy, uncontrolled, careless, verbalistic, intuitive, and muddleheaded.

A main conclusion of Meehl's analysis of the statistical versus clinical debate is also very relevant to our discussion. In his view, quantitative methods are unavoidable in the validation of generalizations. We agree. And since generalizations about what is and is not effective teaching are needed in order to evaluate and improve teaching behaviors, we believe that quantitative methodology is an essential part of research on teaching effectiveness. We recognize that many attempts to demonstrate quantitatively some of the things about effective teaching that we all "know" to be true have not been successful. However, even the fact that it is still difficult to establish through quantitative methods that some individuals are consistently more effective teachers than others should not be a signal to retreat from quantitative work. What seems a far better approach is to define the requirements of research on effective teaching and then determine how quantitative and qualitative research methods can each contribute to such a research program. Qualitative methods can, for example, help to identify different kinds of teaching behaviors whose effectiveness or ineffectiveness can then be validated quantitatively.

We expect that other researchers will share our view that the integration of quantitative and qualitative methodology is essential in solving the very complex problems associated with the identification of effective teaching. The work of Tikunoff and others (1975) at the Far West Laboratory, for instance, has illustrated how qualitative and quantitative methods can be combined to identify potentially potent teacher variables. Smith and Geoffrey (1968), in their excellent book on the urban classroom, have demonstrated the effectiveness of classroom microethnography as a precursor to the construction of theories that can then be verified through more quantitative work.

Having clarified our general approach to this discussion of teaching effectiveness research, we will turn now to more specific issues. We first will outline the requirements of the kind of research that we believe is likely to result in verified generalizations about the primary dimensions along which effective and ineffective teachers vary. Following this discussion, we will provide an example of the kind of research we believe is needed.

Requirements of Research on Teaching Effectiveness

This discussion of research requirements represents a kind of inventory of unfinished business in the area of teaching effectiveness research. We have built this inventory from experience gained over the past five years in our studies of classroom processes and our evaluations of instructional programs (e.g., Cooley and Leinhardt, 1975a; Leinhardt, 1976). Our work has been directed toward the identification of effective instruction and effective programs. Since classroom practices and instructional programs cannot be studied adequately without attending to the individuals who implement them, the observation and measurement of teaching behaviors has been an integral part of our research.

In our view, research to identify effective teaching must meet six requirements. There must be: (1) student outcome measures on which an assessment of effective or ineffective teaching can be made; (2) measures of teaching behavior; (3) measures of variables other than teaching behavior that are known to be related to student outcomes; (4) a model of classroom processes for use in selecting, constructing, and organizing all these measures; (5) procedures for collecting data on these measures; and techniques for identifying those teaching behaviors that influence the desired outcomes.

By enumerating these six requirements, we do not mean to imply that they must be met in sequential order. Some work can be done concurrently; other activities are prerequisite to or dependent upon the satisfaction of certain requirements. In addition, it is unlikely that all six requirements will be totally satisfied in initial attempts to conduct the kind of research that we are proposing. The results of early studies probably will suggest refinements that then can be incorporated into later research efforts.

Student Outcome Measures

A first requirement of a research program on teaching effectiveness is outcome measures that indicate the degree to which learning has taken place. What is needed are student outcomes that are measurable, that theoretical or empirical evidence indicates can be influenced by teaching behavior, and that are valued by those who are to judge teaching effectiveness.

These three criteria for the selection of outcome measures seem reasonable and straightforward, and many will agree that they should be met. Currently, however, there is really only one set of outcomes that can meet these criteria—namely, measures of student achievement. Numerous test batteries and criterion-based tests are available to measure achievement; teaching behavior has been shown to have some effect on achievement; and achievement is regarded as a valued outcome. That is, people believe that higher achievement in school leads to a better career, which leads to a more satisfying life, and so on. Although the relationship between achievement and a better career or more satisfying life clearly needs further clarification, there is ample research evidence to support the notion that achievement when measured at one educational level, is by far the best predictor of academic performance at the next level. Thus, academic achievement is generally regarded by those involved in education as one of several desired student outcomes.

There are, of course, numerous outcomes other than achievement that come to mind when one thinks of possible measures of teaching effectiveness. Self-esteem, citizenship, attitude toward learning and toward oneself, creativity, and psychosocial maturity are examples of outcomes that could be considered for inclusion in studies of teaching effectiveness. Unfortunately, none of these outcomes clearly meets all of the criteria set forth earlier. Additional work needs to be done to develop better techniques to measure these and other noncognitive outcomes, to demonstrate that they can be affected by teaching behaviors, and to establish the value of noncognitive outcomes (i.e., to show that they are causally related to some desired end).

Within the set of available measures of cognitive outcomes, it is possible to distinguish between program-specific measures and program-general measures. Program-specific measures are frequently criterion referenced and are idiosyncratic to the educational program for which they are designed, making cross-program contrasts difficult, if not impossible. Program-general measures are usually norm referenced and either have no components that are idiosyncratic to any one program or attempt to balance these components over an entire test. The specific type of measures used should depend on the aim of the research. If the goal is solely to identify program-specific teaching behaviors (see Siegel and Rosenshine, 1973), then program-specific tests alone are acceptable. If, on the other hand, the goal is to identify teaching behaviors that are effective in a variety of programs, then one must use an assessment procedure that includes more general content to assess student acquisition of academic material.

Measures of Teaching Behavior

A second requirement of research on teaching effectiveness is measures of teaching behavior. There are literally hundreds of such behaviors, ranging from those thought to be important in implementing a specific instructional program (e.g., prescription writing in Individually Prescribed Instruction) to behaviors regarded as critical in most teaching situations (e.g., effective oral expression). All of these variables conceivably could be included in a research effort to identify effective teaching, but doing so obviously is not feasible given the constraints of time, money, and a school's tolerance for classroom data collection. A more reasonable approach is to: (1) identify observable behaviors that empirical (e.g., Rosenshine, 1971) or theoretical evidence (e.g., Cooley and Leinhardt, 1975a, 1975b) has indicated are related to the desired outcome measures; (2) develop procedures for sampling and deriving measures of such behaviors; and (3) reduce the dimensionality of the resulting set of teaching behavior measures through appropriate scaling techniques.

Both program-specific and program-general teaching behaviors should be included in any study of teaching effectiveness. It cannot be assumed that the program-specific behaviors of all teachers using a particular instructional program are the same and thus do not need to be studied. There is ample research evidence that different teachers implement the same program in different ways. This variation in implementation must be observed, measured, and related to outcomes along with program-general teaching behaviors. Otherwise, it will not be possible to establish consistent behavior patterns that are effective in producing the desired outcomes.

The approach to identifying measures of teaching behavior just outlined has not been used by many researchers, and when it has, the results have not been encouraging. Rosenshine (1976), who is a major chronicler of teaching behaviors that make a difference, identified six clusters of variables that research evidence has suggested are related to student achievement: (1) time spent on learning material, (2) content covered, (3) grouped instruction, (4) direct questions on academic material, (5) feedback on academic material, and (6) direct instruction.

Possibly the reason so few behaviors have been found to be related to achievement is that teaching behaviors do not have a significant impact on what students learn. Or it may be that teaching behaviors are so idiosyncratic that it never will be possible to identify general features of effective teaching. However, so little well-designed work has been done in this area that there really is no sound basis for either of these conclusions (cf. Heath and Nielson, 1974). Past research has been plagued by a number of problems, including the variable meanings attached to single labels of behavior and the variety of labels that describe essentially a single behavior. Examples of problematic labels are warmth, questioning, and clarity. This labeling problem is one that can be solved through the development of operational measures of behaviors implied by the labels. More work, both quantitative and qualitative, needs to be done to define behavior, to determine if behaviors thought to be related to student outcomes are, in fact, related, and to identify other behaviors that may affect outcomes.

One measure that has been largely disregarded is a measure of what the teacher knows. Although it can be assumed that a teacher's knowledge, both pedagogical and specific subject matter knowledge, has some bearing on student achievement, there have been few attempts to establish this relationship empirically. A notable exception is the large-scale study by Coleman et al. (1966) in which one of the few "school effects" they did find was the relationship between a simple test of teachers' vocabulary and student achievement. What the teacher knows generally has been overlooked because everyone agrees that knowledge alone is not a sufficient condition for effective teaching; consequently, research has tended to focus on the many other behaviors that are likely to be important, such as how teacher knowledge is used. Measures of any single variable are never sufficient information for assessing teacher effectiveness, since different teachers can have different strengths and weaknesses and still produce similar outcomes. But the study of a particular teaching behavior does provide information relevant to the likelihood that student outcomes will be affected by that behavior. More knowledgeable teachers may, for example, tend to produce more knowledgeable students. Therefore, such information should be included in studies of teaching effectiveness.

Measures of Other Variables

Identifying effective teaching is complicated by the fact that several variables other than direct teaching behaviors are related to student outcomes. Since we believe that studies of effective teaching must take place in the context of studies of effective instruction, we view the identification and measurement of these other variables as another requirement of research on teaching.

At least three major clusters of variables related to instruction need to be considered: (1) initial student differences, (2) the instructional effectiveness of the curriculum being used, and (3) the quantity of schooling provided. Any effort to identify effective teaching must include measures of these variables so that it will be possible to sort out their effects from teaching effects. Doing so is difficult, since teaching behaviors influence both the effectiveness of a curriculum and the quantity of schooling. Various statistical techniques can, however, aid in identifying the unique contributions of each of the major instructional variables, including teaching behaviors, that impact on what students learn.

The *initial student differences* that must be measured are those known to be functionally related to the desired outcomes. If the outcome is student achievement, then students' abilities as they enter a classroom at the beginning of a school year or as they begin a new instructional program will always be the strongest predictors of what they will achieve. Generally, initial abilities can be measured by using alternative forms of the same test used to measure end-of-year or end-of-program achievement.

In our view and that of other researchers (see Cooley and Lohnes, 1976), measures of initial student differences reflect measures of community, home, and peer group influences, unless these influences change dramatically following the assessment of initial differences. Thus, we think that, in general, these environmental influences do not need to be measured directly.

In measuring the *effectiveness of the curriculum* a teacher is using, at least two important aspects must be taken into account. The first has to do with the instructional quality of the curriculum—that is, how well does it teach? The second aspect concerns the degree to which the curriculum content matches what the outcome measures assess. A math program, for example, may do an excellent job of teaching computation skills, but if the outcome measures emphasize skills in solving word problems, the effects of the program on student outcomes will probably be somewhat diminished.

In addition to measures of initial student differences and of curriculum effectiveness, there must be measures of the *quantity of schooling* to which students are exposed. It is obvious that students will tend to learn what they spend time trying to learn and will tend not to learn what they don't spend time trying to learn. Since schools, curricula, and teachers all vary in the time they provide for student learning, some measure of the time students actually spend in instruction must be incorporated into teaching effectiveness studies.

It should be obvious by now that we view the problem of identifying effective teaching as an aspect of the problem of explaining variation in student outcome measures. Only by observing, measuring, and incorporating in the analysis all of the major influences that may impact on desired outcomes will it be possible to identify specific teaching behaviors effective in producing those outcomes. Teaching effectiveness research that ignores other important influences on student outcomes will simply add to the large collection of unreliable, inconsistent results already available.

Model of Classroom Processes

Another requirement in studies to identify effective teaching is a model of classroom processes. Such a model can serve two functions. It can aid in the selection and generation of measures of teaching behavior and other variables that impact on student outcomes, and it can provide a systematic way of combining all these measures into major constructs. Since it is likely that a vast number of measures will be identified, some systematic way of combining them into constructs is needed to facilitate data collection and analysis. Otherwise, the researcher will be left with data from which no clear insights can be gained as to what teaching behaviors are effective in producing the desired outcomes. A model will make it possible to bring some order into what inevitably would be chaos if data collection and analysis proceeded in an unsystematic fashion.

To be most useful, a model should meet several criteria. First, it should be simple; the constructs that it includes should be as few in number and as unambiguous as possible without distorting and oversimplifying the phenomena under study. Second, the model's boundaries and limitations should be explicit. Third, it should be consistent with empirical data and best guesses as to what constitutes effective teaching and instruction. Finally, the kind of information that is generated should be easily interpretable and suggestive of possible additional research, policy relevant actions, and/or refinements in the model itself.

Procedures for Collecting Data

In addition to a model that meets the criteria set forth above, specific procedures are needed for gathering information on student outcomes and the many variables that may impact on these outcomes. Before elaborating on this fifth requirement, we want to make it clear that, in our view, studies of teaching effectiveness should include the collection of data in actual classroom settings. Through classroom research, it will be possible to identify, measure, and relate to student outcomes all or at least many of the variables, including teaching behaviors, that can explain those outcomes. Such is not the case in laboratory research where only a limited number of variables can be investigated at any one time. We do not discount the value of this kind of research. In fact, we believe that laboratory experiments, just like qualitative research methods, can contribute to and draw from quantitative field research and that interaction between the various approaches should be encouraged. However, we view quantitative classroom research as an essential ingredient in studies of teaching effectiveness if one wants to make convincing generalizations about what behaviors will be effective in the classroom.

In defining data collection procedures, there are several important considerations. First, the data must be gathered as unobtrusively as possible. No school official will permit a research study to totally disrupt school operations. Some level of inconvenience will be tolerated, but every effort must be made to ensure it is not exceeded. Second, the procedures must maximize the validity of the information obtained. One way of doing so is by collecting information on the same variables through more than one technique.

Third, procedures must be designed to maximize the accuracy of the data. Since in most situations it will not be feasible to gather information on a daily basis, the procedures must ensure that the information collected accurately represents what occurs when data are not being collected. Another consideration is the need for a permanent record of classroom activities to make it possible to analyze data outside the classroom setting and to re-analyze it using a different statistical approach or different research questions.

Data Analysis Procedures

A sixth requirement of research on teaching effectiveness is data analysis procedures and statistical techniques for identifying those features of teaching behavior that are effective in producing the desired outcomes. A major consideration in defining an appropriate strategy is the unit of analysis. There are, of course, two primary possibilities: the student and the teacher. The fact that each student in a classroom will be treated somewhat differently by the teacher from one day to the next and that some students will receive consistently different treatment suggests that the student should be the unit of analysis. However, it must be kept in mind that teacher effectiveness studies aim to provide information about teachers. Moreover, it is not feasible to collect and analyze data on each student's interactions with the teacher and, at the same time, collect information on a sufficient number of teachers to generalize about teaching behaviors. For these two reasons, we favor using the teacher as the unit of analysis.

A second consideration concerns the number of dimensions needed to represent each of the major variables thought to impact on desired student outcomes. Unless the number of dimensions is small, it will be necessary to collect data on an inordinate number of teachers. For example, at least sixty teachers would be needed to avoid overfitting the data if six dimensions were defined. Although some form of factor analysis is frequently used to reduce dimensionality, this technique has been found unsatisfactory in classroom research (e.g., Leinhardt, 1976; Stallings, 1975). Combining measures of classroom environment into linear combinations because of their patterns of intercorrelation often results in uninterpretable factors. Of course, factor analysis is always most useful when measures have been constructed with that analysis technique in mind. Further, factor analysis has the advantage of minimizing the correlations between factors. However, in classroom research, what has been most useful to date is combining measures that derive from the same construct, where the construct is part of some model of classroom processes that explains desired outcomes.

Once the unit of analysis and the primary dimensions have been established, it is necessary to select appropriate statistical techniques for determining the relative influence of teaching and other variables on student outcomes. Since some of these variables will have nonlinear effects and some will be correlated, a technique is needed that will allow the researcher to sort out a variable's unique effects from the effects that are confounded with those of other variables. It is, of course, quite difficult to argue causal relationships among variables in nonmanipulatory surveys. Everyone knows that correlation cannot prove causality. What is required are analyses that create the strongest valid presumption of causality. Researchers in nonexperimental disciplines such as economics and sociology have been working on this problem for some time and have found techniques such as path analysis to be somewhat useful in their work. However, more satisfactory statistical tools for dealing with the causality problem need to be developed. Toward that end, other regression approaches are being examined.

In making inferences about the relative importance of different variables, it is essential to specify the overall nature of the sample with respect to each variable. Obviously, if all teachers in a sample are using the same approach with respect to a particular variable (e.g., providing the same amount of opportunity for children to learn in each subject matter area), then it will not be possible to determine with that sample the importance of that variable in populations that are heterogeneous as far as that variable is concerned.

An Example of Research on Teaching Effectiveness

In this final section of the paper, we will illustrate the kind of research that we believe can contribute to identifying effective teaching behaviors. Called the Instructional Dimensions Study, this research was designed at the Learning Research and Development Center (Cooley and Leinhardt, 1975b) and is now being conducted with our assistance by Kirschner Associates under contract to the National Institute of Education. It is part of a general examination of compensatory education programs that Congress has directed NIE to undertake. Its primary purpose is to determine the success with which various educational approaches are compensating for children's initial educational disadvantage by meeting their individual needs. An important byproduct of this study should be information on general features of teaching that impact on student outcomes. Our description of this research effort is organized according to the six requirements for research on teaching effectiveness just outlined.

Student outcome measures. The student outcome of primary interest in the Instructional Dimensions Study is achievement in reading and mathematics. Because achievement in these two areas is undoubtedly a valued outcome of education, it is important to understand what produces observed variations in the levels of achievement reached by students in different classrooms.

Reading and math achievement is being measured by the Comprehensive Tests of Basic Skills. Based on such considerations as validity, reliability, quality of format, and cultural fairness, we believe that the CTBS is the best measure of this outcome currently available. In addition, the items in the test seem to reflect a balanced sampling from the various curriculum models being used in schools today. In future studies, better achievement measures and better measures of the variables that influence achievement should be available, partly as a result of research efforts such as the Instructional Dimensions Study.

It is quite important to emphasize that using a standardized test such as the CTBS as an outcome measure in a large-scale research study that takes into account the wide variety of variables that can influence test scores is significantly different from using a standardized test as an administrative device for evaluating the performance of individual teachers. We agree with Glass (1973:53) that "evaluating teachers by measuring their pupils' gains from September to June on commercially available standardized tests is patently invalid and unfair." However, one way to achieve valid and fair assessments of teachers is to: (1) identify through research the many variables that influence student performance on some type of general achievement test; and (2) use those variables that relate to teaching performance as a basis for observing, diagnosing, and improving teaching behavior.

A second outcome of interest is student attitudes toward schooling as assessed by the Survey of School Attitudes. Although it is not clear if this outcome can be measured validly, if it is related to some desired end, or if it can be affected by teaching behaviors, curriculum effectiveness, or any other school-related variable, its inclusion in the study will provide at least some information on its relationship to schooling. By such exploratory work, it may be possible to identify outcomes other than achievement that meet the criteria for outcome measures proposed earlier.

Measures of teaching behavior. The measures of teaching behavior relate to the following: specification of curriculum objectives, matching of students and curriculum, sequencing and pacing of instruction, grouping of students for instruction, quality of the instructional interactions between the teacher and students, amount of time provided for instruction, and noninstructional teacher-student interactions that support and encourage learning. To see how these behaviors are translated into measures, let's examine one set.

A set of six measures is used to assess noninstructional teacher-student interactions that encourage learning. These measures assess the extent to which teachers use praise, the frequency with which they exhibit punishing behavior, and the degree to which they encourage the use of games and contests, student self-evaluation and self-management, and peer tutoring. Theoretical and empirical evidence, including evidence obtained through qualitative research methods, has suggested that all these behaviors are somehow related to student outcomes. Here is an illustration of the kind of fruitful interaction that is possible between quantitative and qualitative methodology; some of the teaching behaviors identified through qualitative methods as potentially important variables in what students learn are now being studied quantitatively.

Measures of other variables. As noted earlier, three major variables other than teaching behaviors also are related to student outcomes: (1) initial student differences, (2) the effectiveness of the curriculum being used, and (3) the quantity of schooling provided. In the Instructional Dimensions Study, initial student differences are being assessed by alternate forms of the CTBS and the Survey of School Attitudes. The measures of curriculum effectiveness relate to the specification of objectives, matching of students and curriculum, sequencing and pacing of instruction, amount of overlap between the curriculum content and what is assessed by the outcome measures, and curriculum motivators that support learning.

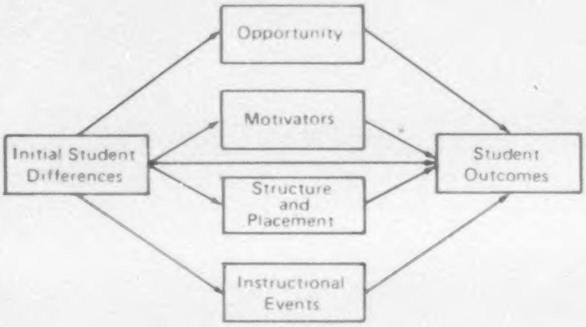
Some of the measures of teaching behavior fall into some of these same categories. For example, some of the measures related to sequencing and pacing are clearly measures of teaching behavior. The extent to which the teacher follows the sequence and creates supplementary materials for improved sequencing are examples. Some of the other measures, such as clarity of the sequence, are primarily measures of curriculum effectiveness. We use the word "primarily" because no curriculum is entirely independent of the teacher who is using it. Thus, curriculum effectiveness measures like clarity of the sequence cannot be considered solely as measures of how well a curriculum teaches. There will always be some confounding of such measures with measures of teaching behavior. However, data analysis techniques can help to sort out the confounding of measures from their unique effects.

Quantity of schooling is being assessed by eleven measures. Some, such as amount of homework assigned, are clearly related to teaching and curriculum differences. Most are not–for example, number of minutes in the school day, number of students in the classroom, and number of adults in the room. Quantity of schooling, therefore, is considered separately even though it, like curriculum effectiveness, may be confounded with other major influences on student outcomes.

Model of classroom processes. The model illustrated in Figure 1 is being used to organize measures of the variables that are assumed to influence student outcomes. This model (Cooley and Leinhardt, 1975a, 1975b; Cooley and Lohnes, 1976) specifies that student outcomes are a function of measures of initial student differences, teaching behavior, curriculum effectiveness, and quantity of schooling.

FIGURE 1

Model Used To Organize Measures in the Instructional Dimensions Study



Measures of the latter three variables are organized according to four constructs: (1) structure and placement, (2) instructional events, (3) opportunity, and (4) motivators. Some examples of the measures included in each of these four constructs are listed below (adapted from Cooley and Leinhardt, 1975b).

Structure and placement. Measures of teaching behavior and curriculum effectiveness related to:

1. Specification of objectives–clarity of objectives, degree to which materials match objectives.
2. Matching of students and curriculum–presence of placement, monitoring, and mastery assessment procedures in curriculum, presence of informal assessment procedures.
3. Sequencing and pacing of instruction–type of sequencing in curriculum, extent to which teacher follows sequence.
4. Grouping for instruction–type of grouping, number in groups.

Instructional events. Teaching behavior and curriculum effectiveness measures concerning:

1. Management information–frequency of management statements, frequency of cognitive management statements.
2. Cognitive teaching to individuals or small groups–frequency of cognitive questions, frequency of child initiated responses.
3. Cognitive teaching to the whole class–frequency of cognitive statements alone, frequency of child responses.
4. Indirect teaching behavior–frequency of personal statements, frequency of extended tutoring time.
5. Quality of teaching techniques–degree to which teacher focuses child's attention, degree to which teacher manages class effectively.

Opportunity. Measures of quantity of schooling, teaching behavior, and curriculum effectiveness dealing with:

1. Amount of time available to learn subject matter–number of minutes in subjects, amount of homework assigned.
2. Curricular overlap–overlap of math materials with criterion test, overlap of reading materials with criterion test.

Motivators. Measures of teaching behavior and curriculum effectiveness related to:

1. Curriculum motivators that support learning–degree of interest of materials, number of modes of instruction.
2. Interpersonal motivators that support learning–degree of use of peer tutoring, degree to which teacher uses praise.

Procedures for collecting data. Data on the measures included in the Instructional Dimensions Study are being collected through four methods. Data on initial student differences and student outcomes are being gathered through the administration of standardized instruments in several hundred grade 1 and grade 3 classrooms. Information on teaching behaviors, curriculum effectiveness, and quantity of schooling was collected in these same classrooms early in the 1976-77 school year and will again be collected in the spring of 1977.

Three techniques are being used to assess these variables: (1) teacher interviews, (2) analysis of curricula by curriculum experts, and (3) videotaping of classroom activities. Teacher interviews are extremely useful in determining specific classroom practices. In general, teachers attempt to provide accurate information, particularly if they do not feel threatened by the questions asked (Leinhardt, 1975). The fact that interviews take place in the classroom encourages teachers to be precise in their responses. Curriculum analysis provides detailed information about the structure and quality of curriculum materials and also helps to cross-validate information gathered from teachers. Videotaping both contributes to the cross-validation of teacher interview data and provides unique information about classroom activities, particularly the interactions between teachers and students. Taping requires fewer highly trained observers than in-class observation, eliminates the possibility of confounding observers with sites, and provides a permanent record of activities that makes it possible to monitor coding accuracy, recode ambiguous results, and re-analyze data later using both qualitative and quantitative research methods.

Data analysis procedures. In describing and analyzing the large amount of information that is being collected, the first task will be to systematically reduce the data. The various measures of teaching behavior, curriculum effectiveness, and quantity of schooling will be reduced to a manageable number of dimensions along the lines of the constructs of the model illustrated in Figure 1.

Data reduction will involve at least six steps: (1) elimination of unusable measures, (2) preliminary correlation and partial correlation analyses within constructs, (3) inspection and reflection of measures, (4) plotting and transformation of data, (5) development of standard scores with unit variance, and (6) the combination of measures to form variables. In combining measures, the procedure will simply be to add related measures after adjusting them to unit variance. This procedure will reduce the data to a manageable number of variables, which then will be combined with measures of initial student differences and student outcomes for data analysis at the classroom level.

Commonality analysis will be the primary technique for analyzing the data. This technique has been proposed by Mood (1971) and others for use in studies such as the Instructional Dimensions Study where the objective is to understand the relative influence of predictors, but where it is not possible to control experimentally the degree of their relationship. Commonality analysis will make it possible to describe the relative effects of initial student differences and other major influences on student outcomes, both in terms of their unique contribution to explaining variation in outcomes and in terms of contributions that are common to two or more of these influences.

In the very beginning of this paper, we noted that our concern is primarily with the research task of establishing general features that characterize effective and ineffective teaching and not with the clinical task of making decisions about the performance of individual teachers. However, the kind of research that we have described can make important contributions to the evaluation of teacher performance. The generalizations about effective and ineffective teaching behaviors that will result from the Instructional Dimensions Study, for example, can serve as one basis upon which observation instruments for use in rating teacher performance can be built. We also expect that the results of the study will suggest fruitful directions for additional research, both quantitative and qualitative, on what constitutes effective teaching.

References

Coleman, J.S., et al. *Equality of Educational Opportunity.* Washington: U.S. Government Printing Office, 1966.

Cooley, W.W., and G. Leinhardt. *The Application of a Model for Investigating Classroom Processes.* Pittsburgh: Learning Research and Development Center, 1975(a).

Cooley, W.W., and G. Leinhardt. *Design for the Individualized Instruction Study: A Study of the Effectiveness of Individualized Instruction in the Teaching of Reading and Mathematics in Compensatory Education Programs.* Pittsburgh: Learning Research and Development Center, 1975(b).

Cooley, W.W., and P.R. Lohnes. *Evaluation Research in Education.* New York: Irvington Publishers, 1976.

Glass, G.V. "Statistical and Measurement Problems in Implementing the Stull Act." In N. L. Gage (ed.). *Mandated Evaluation of Educators: A Conference on California's Stull Act.* Stanford (Calif.): Stanford Center for Research and Development in Teaching, 1973.

Heath, R.W., and M.A. Nielson. "The Research Basis for Performance-Based Teacher Education." *Review of Educational Research* 44(4):463-84, 1974.

Leinhardt, G. "Observation as a Tool in Evaluation of Implementation." *Studies in Educational Evaluation* 1(2):77-94, 1975.

Leinhardt, G. *Program Evaluation: An Empirical Study of Individualized Instruction.* Pittsburgh: Learning Research and Development Center, 1976.

Meehl, P.E. *Clinical vs. Statistical Prediction.* Minneapolis: University of Minnesota Press, 1954.

Mood, A.M. "Partitioning Variance in Multiple Regression Analyses as a Tool for Developing Learning Models." *American Educational Research Journal* 8(2):191-202, 1971.

Rosenshine, B. *Teaching Behaviors and Student Achievement.* London: National Foundation for Educational Research in England and Wales, 1971.

Rosenshine, B. "Classroom Instruction." In N. L. Gage (ed.). *The Psychology of Teaching Methods.* Chicago: National Society for the Study of Education, 1976.

Siegel, M.A., and B. Rosenshine. "Teacher Behavior and Student Achievement in the Bereiter-Englemann Follow-Through Program." Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, February, 1973.

Smith, L.M., and W. Geoffrey. *The Complexities of an Urban Classroom.* New York: Holt, Rinehart and Winston, 1968.

Stallings, J. "Implementation and Child Effects of Teaching Practices in Follow Through Classrooms." *Monographs of the Society for Research in Child Development* 40 (7-8, Serial No. 163), 1975.

Tikunoff, W.J., D.C. Berliner, and R.C. Rist. *Special Study A: An Ethnographic Study of the Forty Classrooms of the Beginning Teacher Evaluation Study Known Sample.* (Technical Rep. No. 75-10-5-A.) San Francisco: Far West Laboratory for Educational Research and Development, 1975.

EFFECTIVE TEACHING: A QUALITATIVE INQUIRY IN AESTHETIC EDUCATION

Louis M. Smith
Washington University

"Effective teaching" is one of several deceptively simple appearing labels, if not conceptions, which abound in theoretical and practical discussions in education. My task, as I perceive it, is to begin a discussion based upon qualitative methods of educational inquiry. Since most of my recent work has been in the domain of aesthetic education,[1] it seems appropriate to begin there and gradually expand the generalizability of the discussion to other curricular areas. In part, the underlying logic in selecting aesthetic education as a research area was that it is complex enough that if solutions could be found there, other areas would be susceptible to analysis.

To anticipate the detailed argument and set the direction for its development, let me present the overall conclusion–effective teaching is a complex valuational/theoretical/empirical judgment. By this I mean that the process of coming to a judgment that a teacher is effective or ineffective requires, unavoidably: (1) taking a stand among several values which may be in conflict, (2) taking a stand on a number of conceptional definitions and theoretical propositions which are only one of several possible ways to construe the domain of teaching and learning, and (3) taking a stand on conflicting empirical evidence both in general and in regard to the particular teacher and learning situation. If all this be true, and I hope to make a case that it is, it has major implications for the conduct of research and practice in teaching. Neglecting any one of these three domains produces judgments less than adequate to the task at hand.

One style of presentation in our qualitative research has been to involve the reader directly in the field situation through excepts from field notes and brief accounts we call vignettes. Two such vignettes, adapted from Smith and Schumacher (1972), follow. From them we will move inductively to hunches, insights, analytic conceptions, hypotheses, and theories.

Vignette 1: Making Music

Imagine you are observing a class using a new curriculum, the Aesthetic Education Program (AEP). You take some notes and write a brief vignette.

This morning at 10:25 a.m. I observed my first class using the Meter package.[2] For a musical illiterate, it proved to be a fascinating experience. The twenty-eight children, second and third graders in an open environment class, were grouped in two semicircles. Center stage was shared by Chart #4, the phonograph, and two children, each with what looked like a homemade drum head. The teacher, who seemed comfortable with the materials, explained "accents" as "louder or stronger" beats and indicated that the children should make a fist for the hard beats and use an open hand for the soft taps. Her directions blended with explanations as she indicated "bar lines," "measures," and "duple" and "triple" meter. The two children who were up front had little trouble reading the music and performing as musicians. This activity was rotated through several pairs of children, each of whom selected his successor. Spliced into the activity was a total group performance. The children clapped the several lines of music in duple and triple meter with appropriate accents. Throughout, participation and involvement were high. The facial expressions were of pleasure. The teacher made almost no comments of a disciplinary sort.

The teacher flowed in and out of the lesson in what might be called "goal facilitation interventions" (Solomon, 1971); when some problem hindered accomplishment, she found a way to move in, momentarily help, and move out. The best example occurred with a child or two who couldn't use the drum head. As though teaching a psychomotor skill, the teacher reached around the child, held the drum and the child's hand, and started the appropriate duple or triple meter. When the child caught on, he carried on alone. Later illustrations occurred in the total group clapping when the teacher would clap in exaggerated fashion, particularly with each new line and new beat. In the middle section of activities, she went from table to table where children were having difficulty. As she said, "Listen!" she would tap the beats and accents on the table with exaggerated, obvious motions. As the children understood, even momentarily, she moved on.

The middle part of the day's lesson was listening to Activity 7, a record, and writing an answer on Response Sheet 3. The children had difficulty following the directions, hearing the meter, and getting responses recorded. The teacher (and the principal who was visiting) moved about to help as indicated earlier. The kids seemed puzzled; their faces and actions did not reflect clarity, they looked at each other's papers, they raised their hands for the teacher. Progressively, more playing with pencils, more reading of library books, and more chattering occurred. Concurrently through this twenty minutes, more teacher comments, "shushes," and "sit right in your seat" directions appeared.

The final fifteen minutes was spent in a total class review—"Go over the materials so you understand"—of the record and identification of the meter and noting of accents. This turned out to be a "mild disaster." The teacher drew the meter charts on the blackboard. She tried to stay ahead by alternating between two boards, one on the south wall and one on the west wall. She had pupils go to the board and indicate meter and accent for each. They had problems. The teacher's comments, "Listen carefully boys and girls; most of you aren't listening," seemed both accurate and necessary. She did the last part "once again," over the growing distraction and resistance of the children, for she was concerned that they understand. Her last comment was, "I think we'll have to do it over again. Some of you haven't got it yet." At 11:13 they started to set up for an ETV science lesson about the moon.

As I reflected on this and similar episodes, ideas arose suggesting several lines of analysis.

1. The lesson had several discernable common sense parts to it. Quantification requires specification of the units (Smith and Brock, 1970).
2. The first part seemed to reflect a recreator or performer role. More precisely, the children—individually and collectively—read the chart, clapped or beat time on the drum, and "talked music." They were on the adient end of an involvement continuum—toward joy.
3. Parts #2 and #3 seemed more in an appreciator/listener/audience role. Affect moved toward the non-involved and avoidant end of the continuum.
4. The causes of the events are another set of issues—the lesson was too long, the music is too complicated, the teacher knows music, the open environment is congruent with performing but not appreciating, etc.
5. For evaluative purposes, any product analysis—e.g., response sheet #3—is hopelessly contaminated with teacher help, principal help, peer help (willing and unwilling), brevity of items, audibility of record.
6. Further insight into the degree of implementation problem might be phrased as "implementation within implementation." In effect, a curriculum diffusion model is being implemented, a curriculum is being implemented, and finally a lesson is being implemented. Analytically, the same data may yield very different implications depending on which level of analysis one works within. Teacher effectiveness assessment becomes intertwined with these issues.

Vignette 2: Which Objectives When?

In recent years, we have come to believe that much of teaching can be viewed from the perspective of dilemmas a teacher faces.[3] Observing a substitute teacher using *Dramatic Plot*[4] for only the second time, the relevance of this perspective for AEP materials arose. The children have been busy some forty minutes and are culminating the activity by writing the plot they have created. The field notes capture the tenor of that episode.

> Mrs. Wilson comes over with a paper containing capitalization, punctuation, and spelling problems. She's aghast. She also talks of parental concern over need for rapid knowledge of the multiplication tables. The parents were also concerned with the achievement test results that showed the children were a little low in punctuation and usage. That capsules the problems and dilemmas of the teacher very well.
>
> My reactions to her were:
>
> 1. Teachers vary.
> 2. Some accent imaginative stories here.
> 3. And later accent spelling, usage, and punctuation.
> 4. Need to be clear in one's own mind about the objectives, aims, and goals of this activity.
> 5. And when other skills and goals will be accented and taught.
> 6. Possibly taking these very papers and using them as a basis for a specific and more classical language arts lesson.
> 7. Trying to do both at the same time may kill the imaginative part, etc.

We gave that a summary label in the field notes: "Vignette of the dilemmas of a traditional school marm." Upon reflection, the episode seems to encompass a host of dilemmas if the packages are to be used effectively, if teachers are to be trained appropriately, and if diffusion is to occur rapidly and easily.

Substantively, the question of which objectives when seems closely related to a teacher's need for a clear conception of the program and to the issue of the program's relationship and articulation with the overall curriculum. Few advocate omitting traditional language arts goals, but few speak clearly to the theoretical and practical teaching problems in developing AEP harmony with these other curricula.

Problems for a Conception of Effective Teaching

These brief vignettes and interpretive comments are only a few of many from our notes and reports. They raise a number of questions and problems when I try to interpret them with the beginning definition of "effective" and "teaching." Our problem now is to unpack the label "effective teaching" in the context and specificity of our illustrations. In ordinary language, effective is defined by Webster as "producing a decided, decisive or desired result" and teaching is defined as "to make to know how; to direct...guide the studies of; to impart the knowledge of; to make aware by information, experience or the like." In the meanings of ordinary language, then, effective teaching becomes producing–via making, directing, imparting knowledge, and making aware by information or experience–a decided, decisive, or desired result, a knowing how, a knowledge of, an awareness. In more sophisticated form, Scheffler (1971:121) has defined teaching with a critical extension or two:

> ...an activity aimed at the achievement of learning, and practiced in such manner as to respect the student's intellectual integrity and capacity for independent judgment. Such a characterization is important for at least two reasons: First, it brings out the intentional nature of teaching, the fact that teaching is a distinctively patterned sequence of behavioral steps executed by the teacher. Secondly, it differentiates the activity of teaching from such other activities as propaganda, conditioning, suggestion, and indoctrination, which are aimed at modifying the person but strive at all costs to avoid a genuine engagement of his judgement on underlying issues.

Similar positions have been taken by Green (1971) and Peters (1965). Such conceptions are useful in distinguishing teaching from other influence processes and in highlighting the child's developing "mind," humaneness, and autonomy.

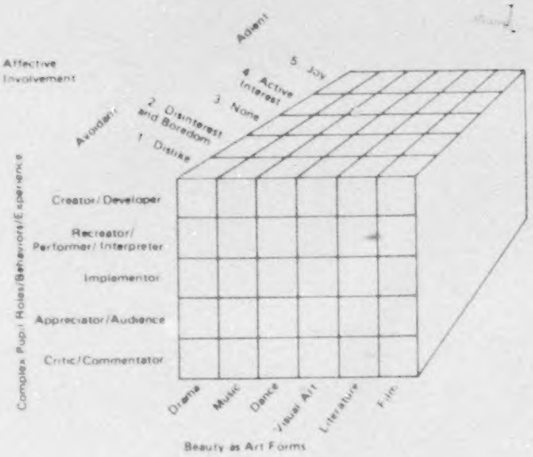
Defining and delimiting the curricular domain. Based on these definitions, one of the most basic problems in making a judgment about effective teaching is in defining the curriculum domain. This problem is particularly acute in aesthetic education. In my judgment, there are multiple and only partially overlapping conceptions. At a general level, Broudy (1972) speaks of aesthetic education as enlightened cherishing; Barkan et al. (1970) of an introduction to aesthetic experiences, those which are satisfying in themselves; and Madeja (1971) of a loosely organized area of study, comparable to social studies or language arts. One task we set ourselves was to develop from our data a preliminary model of aesthetic education. As the vignettes illustrate, we found empirically that teachers were involving children in multiple art forms (music, drama, dance, graphic arts, poetry) and multiple experiences (producing, performing, implementing, appreciating, critiquing). These conceptions culminated in the model represented in Figure 1 and Table 1.

This point of view of the domain of aesthetic education has been instrumental in later theoretical and empirical analyses (Smith, 1974a, 1974b, 1975; Smith and Greenberg, in process) and has been useful in the practical problems of evaluation and of teaching in teacher education programs.

In short, teaching has a conceptual base in a curriculum domain. One kind of theoretical analysis in any discussion of effective teaching is the structure of the domain being taught (Bruner, 1960; Hirst, 1974). In a sense, it functions much as a specification table does for measurement and evaluation; it specifies content decisions. For our purpose, it structures the territory which is relevant for teaching decisions: What is to be included? What experiences–how organized and sequenced–are children to have? What constraints of time, space, resources must be worked around?

FIGURE 1

A Model of Aesthetic Education



An important question is how much these decisions lie with the teacher or are directives from the larger system. In the Washington School (Smith and Geoffrey, 1968) a traditional inner city school, the domains were spelled out in minutes per week, but teachers made their own decisions and deviated markedly from official guidelines. This indicates an even broader set of issues. Is a teacher who defines the curriculum for inner city children as "the three R's, the basics" and ignores science, literature, and the arts an effective teacher? In my view, such a question cannot be addressed without a conception of the overall curriculum domain and of the specific curriculum area. In other settings, issues in defining the structure of the domain are complicated greatly. For instance, at Kensington, an innovative suburban school, part of the innovation was giving teachers and children responsibilities for openly addressing and making these decisions (Smith and Keith, 1971).

A judgment of effective teaching must take into account these varying conceptions of the domain. As teachers hold different vies of the domain, different events will occur in the classroom and the teaching will be subject to different judgments of effectiveness.

Establishing priorities. The conceptual problem of domain quickly becomes a value problem. To make a judgment about effective teaching, priorities need to be set. For aesthetic education, we have identified five experiences the child might have–creating, performing, implementing, appreciating, and critiquing. Are

TABLE 1

Further Specification of Roles and Art Forms

Pupil Roles/ Behaviors/ Experiences	Art Forms					
	Theater	Music	Dance	Visual	Literature	Films
Creator	Playwright	Composer	Choreographer	Artist	Writer	Playwright
Recreator/ Performer	Actor	Musician	Dancer	Copyist	Oral Reader	Actor
Implementor	Producer Director Stagehand Designer	Conductor Accompanist	Accompanist	Exhibitor	Editor Librarian	Director Producer Cameraman
Appreciator	Playgoer	Concertgoer	Ballet Enthusiast	Art Patron	Bookworm	Movie Buff
Critic	Drama Critic	Music Critic	Dance Critic	Art Critic	Book Reviewer	Film Critic

all five equally important? Should children spend 20 percent of their time in each area, or is one more important than another?

To make a judgment of effective teaching requires some agreement on priorities in the subparts of the curricular domain. Several groups contend to influence such curricular priorities, including patrons, members of the profession, and students themselves. Just as much of beauty is in the eye of the beholder so, too, much of what is good or important in teaching lies in the priorities of a culturally pluralistic society. To ignore the possible conflicts in objectives, goals, and priorities is to miss a key issue in assessing teacher effectiveness.

An important example are the materials in the Aesthetic Education Program which the developers view as curriculum *resources*, not as *a* curriculum. Each community, school, and teacher is to select packages and activities according to individual priorities. As our second vignette illustrates, an outsider who finds a teacher using *Creating Dramatic Plot* or *Creating Word Pictures*[5] as vehicles for teaching paragraphing, punctuation, and parts of speech cannot argue that this is ineffective teaching unless he knows the teacher's objectives and priorities. Nor can one be too harsh on the curriculum writer who feels that the "real" intent of the materials has been bastardized by an "unknowing" teacher. While discussions of such conflicts are not rare in curriculum development organizations, the full implications of the autonomy expected of teachers and local schools seldom are pushed to their theoretical and practical conclusions.

Pupil responsiveness, involvement, and participation. One of the most troublesome problems I've had in conceptualizing the aesthetic education program has been the role of affect or emotion. The disparate perspectives included items such as: (1) the essence of the aesthetic reaction is emotion; (2) the essence of the expressive is the emotional, the affective; (3) understanding emotions in the human condition is a major function of aesthetics; (4) the children should come to like some/all kinds of aesthetic experiences; (5) the teacher should help the children to approach the material and experiences "positively." Now, in long retrospect, I believe that the dimension of affect in Figure 1 refers to the fifth perspective, the child's approach to the experiences in the art forms. Its importance lies in its explication of that point made in common by Scheffler (1971) and Peters (1965) in their conceptions of teaching—the child coming to the learning of his own accord. This is a requisite part of the definition of teaching, necessary to distinguish it from propaganda or indoctrination and to link teaching with the end state of autonomy.

We have not dealt with pupil responsiveness as an entering condition, although it seems obvious that some children and some classes are much less docile, in the sense of responsiveness to teaching. Some of the conventional wisdom regarding social class and teaching seems to fall here. The influence of compulsory education for children 7-16 years and of required courses seems important. The interaction of the novelty and level of difficulty of materials on pupil responsiveness seems to be an important process phenomenon. Finally, the effects of teacher skills, as indicated by Kounin's (1970) high multiple R's relating teacher actions to work involvement, suggest important implications of pupil responsiveness as a dependent variable to teacher efforts.

> Hyperactive or excitable classes contrast with one we labeled, "the lethargic class." Labeling or typing classes has its hazards. If the labels reflect critical dimensions, then the process can help produce the careful thinking or the muted cues that contribute to successful teaching. One of our groups was described as "below average in ability and difficult to arouse emotionally." The teacher seemed to be doing several things as she taught a lesson in Creating Word Pictures. First, she let them organize themselves into groups of fours which took about six minutes. As she said "If I let them find themselves it works easier." Second, she had monitors come up for books, cards, and large layout sheets. Third, she told them to turn

to page 28. Fourth, she presented the guiding ideas with a reference back to "grasshopper wallpaper" and "wallpaper grasshopper." Fifth, she read the first page or two to the children. Sixth, she made sure everyone could find the "spot with the spider." Seventh, she finished reading. And finally, the teacher started to move around among the children giving help and suggestions. The observer noted "Through all this the kids are quiet, attentive, interested."

Insofar as pupil responsiveness is in part an "entering condition," in part susceptible to "momentary teacher influence," and in part a "long-term objective," a judgment of effective teaching has major complications.

The Problems of Means

To this point, we have dealt primarily with the problems of ends, goals, and purposes. Equally difficult problems lie in the analysis of means and instrumentalities. Among the more critical are (1) the functional equivalents problem, (2) multiple conceptions of means–teacher behaviors, teacher roles, teacher-pupil interactions, pupil behaviors, (3) the technical difficulty of various means, and (4) long- and short-term emphases, the instructional versus the motivational dilemma.

Functional equivalents in teaching. In my judgment, the most important unsolved "means" issue in a theory of effective teaching is one I would label the functional equivalents problem. The term is from Merton (1957) who uses it to indicate that quite different overt or manifest items in a group, organization, or community may lead to the same ends or consequences. In essence, they are functionally equivalent. Research which treats these "differences" as differences will show considerable no differences in empirical tests. In most complex field research, whether quantitative or qualitative, naturalistic or systematic, experimental or correlational, innumerable events are beyond researcher control. In the practical identification of the individual effective teacher in the on-going naturalistic situation, the problem is further complicated.

Illustratively, some major categories of functional equivalents I would suggest are teacher action, materials and textbook exercises, parental behavior, the child's cumulating personality structures and processes (general traits, abilities, mathemagenic behavior), and community and organizational items. Such functional equivalents problems arise with AEP. For instance, one of the more unusual patterns of interaction in using the *Creating Dramatic Plot* materials occurred late in the year in what was ostensibly a review lesson. The field notes capture the specifics.

> In several of the plots, there are three characters and 3 kids. Each becomes one of the characters and they alternate who does what. This makes the whole effect much more personalized. The kids take on the roles right from the start. Later apparently they'll act it out–skits. This kind of improvisation in the creative construction phase is very different from the more distal aspects in other groups. Also it involves the kids in performing during creation. (Obs: Is that easier? Eliminates the need for rehearsal?)

On their own initiative, the children have novelly reconstrued the task, broadening it significantly from the original intentions of the developers.[6] Is the teacher to be credited or blamed? Is the judgment of this episode one of more or less effective teaching?

We ran into other events which pertain to the functional equivalents problem. An early memo commented:

> Recently, during one of my visits to the site, I encountered a set of experiences about which I've heard little talk and seen little written by AEP personnel. It developed this way. Upon entering the school, I was face to face with a group of older elementary kids who were square dancing. Later I was told they were getting ready for a Spring festival, which would also include vocal music and guitar playing. After watching that for few minutes I found my way to a cup of coffee in the teachers lounge where several teachers were good humoredly teasing the itinerant art teacher about her morning activity. She was sewing together some nine small pot-holder-size weavings of the children into one large colorful wall hanging. As they were instructing her in arranging the multicolored pieces to eliminate clashes among the pinks, they chided her that the ragged back was as pretty as the front but that *they* wouldn't want it hanging in *their* rooms. In keeping with the tenor of the interchange, I was duly hesitant about taking it either. Later, after watching a creative lesson in which the meter package activity was improvised into an orchestra–drums, rhythm sticks, clapping, and a leader who used the teacher's silk head scarf as a director's wand, I walked out to recess with the teacher. Across the field the junior high marching band was practicing for a crippled children's walkathon. The teacher was soon telling me about the high school band which was good enough to win prizes at the Mardi Gras and in a three-state competition.
>
> The question all of this raises is where does AEP fit into a culture like this? Is there something that might be called Participatory Folk Art? Issues of cultural pluralism, disappearing rural localist traditions, the Deweyian concern for separatism of fine arts from practical arts come to mind. Those of you with broader backgrounds in culture and the arts will see other more subtle implications. On the surface, it seems worthy of attention.

Now, after more careful review of the data, the label "participatory folk art" seems too pretentious for most of our settings. The schools do have programs in music, art, and in some instances drama. Considerable direct and incidental instruction occurs. In the late spring, we ran into annual music concerts, festivals, and programs in school after school. One school had an extended arts program. It had the traditional spring assemblies. The halls housed a collection of original paintings and sculpture. Two weeks in February were set aside for a total restructuring of the school program; teachers and local citizens who are specialists in puppetry, dance, improvisational drama, crafts, and so forth taught special units which children elected according to interest. There was an optional summer extension of these experiences. At the time, we were perplexed about how to handle such phenomena in our evaluation reports. The implications both theoretically and practically for problems in the definition and identification of effective teaching seem equally great.

Recently, the provocative work of Toffler and McLean has suggested directions that bear directly on this aspect of functional equivalence. In a beautiful piece of conditional inference about the art of measuring the arts, Toffler (1974:63) comments that while there is no widespread, much less universal, definition of quality of arts in a society, one can make a start.

> Imagine a society whose cultural output was (1) copious, (2) richly varied, (3) technically outstanding, and (4) marked by many works of excellence. Imagine further that a significant portion of this output represented (5) contemporary creative work, as distinguished from performances or reproductions of the finest works of the past. Assume that much of this output was also (6) of such high complexity that it required (7) a considerably sophisticated audience. Now imagine that a large and sophisticated audience did exist, and moreover, that it was (8) growing in size and that it was (9) highly committed to cultural activities. Imagine there to be (10) a vast amateur movement providing a training ground for both artists and audience. And assume further that the institutions of art, such as museums, theaters, and arts centers, were (11) geographically decentralized, and increasing in number, size, and the efficiency with which they disseminated the work of artists to the public. Suppose that artists in this society were (12) held in high esteem by the public, (13) well remunerated, and that (14) among them were men of undoubted genius. Finally, imagine that the artistic products of this society were (15) consistently applauded in other countries around the globe.
>
> Looking at such a society, might one not draw certain conclusions about its cultural life? Might one not be justified in referring to its high quality?

Toffler then explicates the kind of social indicators that might be counted and quantified relevant to his hypothesized community. McLean (1975) revised the position and applied it to "judging the quality of a school as a place where the arts might thrive." The revision moved toward developing codable systematic observation schedules of variables such as quantity, diversity, excellence, originality, and vitality. They seem directly applicable to the two schools heavily involved in participatory arts and crafts.

An extension of this work to the functional equivalents problem suggests that an elementary classroom might be observed in the same terms. In our own research, we have not gone beyond discussions in seminar. The kind of model that I have in mind can be seen in the contrast between the classroom reported in Richardson's *In the Early World* (1964) and our report in *Complexities* (Smith and Geoffrey, 1968). If one assumes each is a veridical statement of classroom events, then a simple content analysis using the Toffler/McLean type dimensions would suggest the nature of the major differences in the classrooms. Further, if one assumes that the teacher is the major determinant of classroom events, the beginnings of judgments can be made regarding effective teaching of aesthetic education.

Briefly, then, when very diverse actions on the part of teachers, children, materials, parents, and communities accomplish the same desired ends and when they appear in unknown amounts and interrelationships, the judgments about effective teaching become very difficult.

Multiple conceptions of means. Theoretical eclecticism makes things easy when one faces difficult, disjunctive practical problems which seem to come from different directions and to be important for different audiences. But it is maddening for the compulsive, parsimonious, and orderly theory builder. The problem is simple to state, but devastatingly difficult to solve. Educators who talk about teaching draw from multiple, partially overlapping, and often conflicting general social science theories. For example, we speculated in our notes about teachers who seemed to find the AEP materials so isomorphic or congruent with their *teaching style* that they moved with the materials as though they were born to them, as though they had a *tacit understanding* of the curriculum.

> Another one of the experiments that needs to be run is to take a teacher like Mrs. Johnson and have her teach the AEP stuff and also have her teach comparable groups with the regular stuff. That would be an interesting part of the degree of implementation problem; it would also raise some real questions about what happens when the teachers with the enthusiasm and the *flair* go at the problem with any set of materials. Maybe partly what I am asking is to scale teachers in terms of something that might be called flair and watch them take on different kinds of materials such as AEP and such as the Silver Burdette or other materials. In a sense then, one could get middle level teachers on flair and non-flair type teachers. That would give you an interesting interaction with the materials themselves.[7]

Flair (which Webster defines as a natural talent or ability) or *style* or even such high sounding words as *tacit understanding* are labels which slip into discussions of teaching. They compete with more recent and generic labels and theoretical concepts such as teacher *action* and teacher *behavior*. Historically, these vie with the classical statements of dominative and integrative teacher *personalities*, teacher centered and pupil centered social emotional classroom *climate*, direct and indirect *influence* or *interaction process*, autocratic and democratic leadership *style* and *roles*. The list seems unending.

Such a mix of operational definitions, conceptual labels, theoretical structures, and metatheoretical positions at the heart of the empirical research on teaching may be research vigor, but it also may be practical chaos. Even the more comprehensive summaries (e.g., Dunkin and Biddle, 1974, and Rosenshine, 1971) have not taken on the needed theoretical codification and synthesis. In the Aesthetic Education Program, thinking remains at the eclectic level.

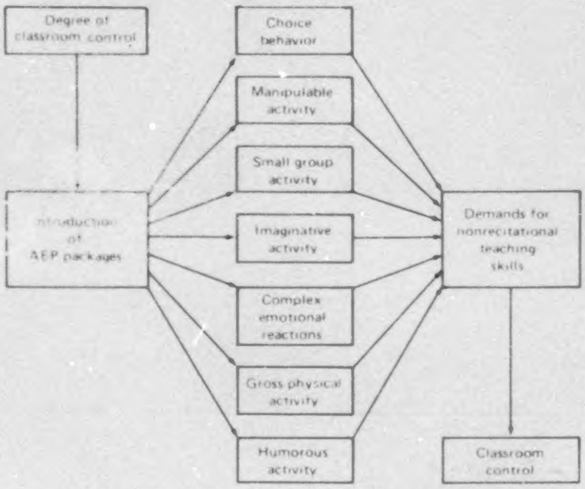
Short-term vs. long-term perspectives. One of the classic dilemmas over which educators divide is the relative importance of what we have been calling the instructional versus the motivational dilemma. To overdraw the extremes, it concerns teaching strategies which stress the child's interest and motivation, "turning him on," versus those which, again in the extreme, stress careful presentation of information and concepts by the teacher. For instance, *Creating Dramatic Plot* introduces such key conceptual elements as character, setting, incident, conflict, crisis, and resolution. The developers intended primarily to excite children about creating dramatic plots and only incidentally to teach the specific concepts. But we found a number of teachers who spent considerable time presenting, defining, illustrating, and then checking on the children's knowledge of the concepts before "playing the dramatic plot game."

Our experience suggests that teachers vary quite dramatically in the degree to which they emphasize one gambit or the other. The linkages of these gambits to different objectives held by the teachers and to the long-range achievement of the children are not clear.[8]

Level of difficulty issue. If we assume that a didactic recitational lesson, in the best sense of Ausubel's (1963) expository teaching and reception learning, is an easier task than a multi-group cooperative activity in drama stressing creative writing, performance, and criticism, then we have a further complexity in conceiving and identifying effective teaching. I'm reminded of competitive diving in which the dives are scored in terms of the diver's performance and the difficulty level of the dive. For instance, dives are graded in terms of level of difficulty; the one-and-a-half gainer with a half-twist in a layout position versus a forward one-and-a-half in a tuck position. The need for such a "difficulty level" for teaching has been apparent with the AEP materials. Most of these materials involve what we have called "teaching a nonrecitational curriculum" and "arousing complex emotions and expressive behavior in the classroom." Both of these "high difficulty" tasks have implications for classroom control—as antecedent and consequence and for multiple kinds of learning. In analyzing the extended pilot trials, we developed the model in Figure 2. As elements of pupil choice, imaginative behavior, emotional reactions, group activities, physical movement, and manipulative activities increase, the demands for nonrecitational teaching skills also increase. Without those skills, control problems can result.

FIGURE 2

AEP Curriculum and Classroom Control



The line between "humor" and "silliness" for example, is very fine. The AEP creative materials have an irrevocable tie with humor. On occasion, with classes with minimal control or with pupils who had personal problems, the materials would set them off. The noise level would increase, the excitement would become contagious, and soon the pupils' behavior would lose any kind of focus intended by the package, leaving the teacher to struggle with the chaos.

Another illustration, drawn from a very brief observation of an activity involving emotions in acting (the *Creative Characterization* package), suggests the array of issues in expressive behavior in the classroom. The children were engaged in improvisational sessions built around role playing charcters on a picnic—Raef, representing fear; Egar, anger; and Har, happy. The field note fragments comment.

1. As the children are thoroughly in their roles the teacher comments, "All right. There is a fine line between staying in the character and acting silly. Don't overdo it."
2. Two children commenting to each other, sotto voce, said, "Billy's acting is not his real self; Joey's acting his real self."
3. In describing and critiquing Joey's performance of Egar, the children indicate, "Everything was good, powerful voice, feeling of anger building up." (Obs: Kids are worked up; difficult to control selves. Once again expressiveness running loose.)
4. Later, a boy who had been playing Raef sought the teacher's attention. He did not get it. He engaged in a variety of attention getting behaviors. (Obs: Boy who was Raef is ticked off. He seems to want commentary on his acting. He also seems like a problem child. Her move to the next activity is a very bad move on above grounds.)

Our vignette suggests one or two simple hypotheses. In contrast to textbook learning, role learning, role playing or dramatic improvisation is more apt to generate emotional and attitudinal reactions. Second, turning off the emotional reaction is not as simple as shifting from arithmetic to spelling. The problem was very real with the boy playing Raef, who was so "jazzed up" by the activity that he could not settle down. Judging teachers who try more difficult teaching tasks and who may have more moderate "success" than teachers who attempt less difficult strategies constitutes a very difficult means problem in judging effective teaching.

Finally, the difficulty level problem intertwines with the priority problem, as when one finds some of the children without necessary skills in cooperative relations and the control of expressive behavior and creative action. At this point, should the teacher take the time to work on these skills at the expense of other kinds of learning? Judges who value "democratic pupil relationships," "group problem solving," and "creative expression" will rate teachers differently from judges for whom these are not important criteria.

TABLE 2

Validity of Participant Observation:
A Multimethod, Multiperson, Multisituation,
and Multivariable Matrix

1. **Methods**
    - 1.1 Observation
    - 1.2 Informal interviews
    - 1.3 Documents: lesson materials, computer print-outs, et cetera
2. **Persons**
    - 2.1 Pupils
    - 2.2 Cooperating teacher
    - 2.3 Principals
    - 2.4 Other teachers
    - 2.5 Multiple incumbents of multiple positions in multiple organizations
3. **Situations**
    - 3.1 Pupils at terminals
    - 3.2 Classroom teaching: announced and unannounced visits
    - 3.3 Multiple parts of the curriculum—in addition to arithmetic
    - 3.4 Multiple schools
    - 3.5 Multiple organizations
    - 3.6 Multiple parts of the country
4. **Variables**
    - 4.1 Individual: schemas, traits, motives
    - 4.2 Group: classroom interaction, activity, sentiments
    - 4.3 Organizational: schools, universities, R&D, Title III

Source: Smith and Pohland, 1976:48

Strategies and Tactics in Generating, Combining, and Reporting Data

In this essay, as in most of our recent work, our data strategies and tactics have involved naturalistic observation producing qualitative data which have been combined rationally but subjectively and reported in vignettes combining description and analysis. As Meehl (1954) indicated so clearly, it is possible to distinguish between kinds of data—qualitative (HSR) and quantitative (ACE score)—and method of combining data—qualitatively (counselor judgment) and quantitatively (regression equations). His argument can be extended to reporting results—vignettes vs. tables. In addition, we have argued that the phase of the research process—generation of concepts, hypotheses, and theory vs. verification/falsification phases—can be related to qualitative and quantitative strategies. Finally, one might argue the nature of the decisions and the economy of effort involved, for some decisions are easier by rule and others by judgment. Briefly, I would take up several of these points.

Building upon Campbell and Fiske's (1959) conception of a multitrait-multimethod matrix approach to construct validity of test data, we constructed a multimethod, multiperson, multisituation, multivariable matrix of data from our computer assisted instruction study. Table 2 is from that report; the categories of data from aesthetic education are comparable.

Presumably, the identification and measurement of effective teaching would require a similar variety of data. Some of these could be quantitative, some qualitative, some from tests and some from interviews, some from systematic observation schedules, some from ratings, and some from qualitative notes. The key issue is that they be valid. Formal organizations seldom give true pictures of their internal operations to multiple, sometimes unknown, and often hostile but relevant audiences in their environment. In our 1971 study of the Kensington School, we came to call this view of the formal doctrine which was presented to the public "the facade." It is my belief that this is a very real and very general phenomenon in education and that it is often ignored or underestimated in otherwise very sophisticated technical analyses in education.

Data can be combined in various ways for various purposes. Descriptively, we have tried to write meaningful narratives which tell the story of intentions, beliefs, actions, and human relationships in a valid, interesting, meaningful, and accurate fashion. We find parallels in the work of the investigative journalist and the descriptive historian. Theoretically, our usual procedure has been a process of concept formation and hypothesis formation in the noting of similarities and differences in episodes of events recorded in the notes. Becker (1958) speaks of selection of problems, concepts, and indices; checking the frequency and distribution of phenomena; construction of social system models; and final analysis. Denzin (1970) uses the label "triangulation," the focusing and combination of multiple methodologies on specific problems and issues. Presumably content analysis, cross tabulations, and multiple regressions are appropriate alternatives for the combining of some data (qualitative and quantitative) for some purposes.

The presentation of our data typically has been in prose accounts—the brief vignette to the long book. The longer accounts have included pictorial models and miniature theories as we have understood these in the work of Merton (1957), Zetterberg (1965), and March and Simon (1958). While our colleagues and students have argued pro and con about such devices, we have found them powerful in facilitating and clarifying our thinking. They also are potent in critically analyzing the "if-then" propositional thinking in the discussions of others. For some other purposes and with other data, tables, graphs, profiles, correlations, and significance tests have an important place.

In recent years, we have been training teachers and administrators in participant observer methods, which we think can be a very powerful, practical set of procedures for the analysis of practical educational issues. The degree to which these methods can handle the problems of effective teaching as a valuational/theoretical/empirical judgment seems open to investigation. The move to assessing a particular teacher's effectiveness would be an important test of such possibilities.

Toward Some Specific Procedures: A Personal Position

At one level, the general thesis and conclusion has been stated simply: effective teaching is a complex theoretical/valuational/empirical judgment. To conceive of it as less than this or different from this is to court a series of potential problems in the discussion of teaching. In route to establishing this conclusion, most of the data, analysis, and argument arose in qualitative studies of teaching and learning in an innovative aesthetic education curriculum. The methodology underlying this inquiry has been the qualitative stance known as participant observation, classroom microethnography, or anthropological method.

One problem with analysis is that it can lead to visions of complexities which in turn can impede action. To sidestep such a possibility, I'd like to sketch briefly a personal position on the valuational/theoretical/empirical judgment which might have some generalizable, if debatable, aspects. Essentially, I'm arguing for: (1) a defensible conception of the program and its priorities exhibited in the teacher's actions; (2) a quality of improvisation in the teacher's behavior; (3) a responsiveness to pupils' suggestions; (4) an involvement and participation by students; and (5) a varied set of changes in pupil personality which accent the multiplicity of possible goals and experiences suggested in Figure 1 and Table 1. My comments will focus on several points which have received less attention in the earlier discussion.

The thread of creativity: a priority. Perhaps I've been overly persuaded by Beittel's (1972) analysis of the making of art and the possibilities this holds for personality development, but for me the experiences/roles/behaviors that lie in the creator/developer row of the models in Figure 1 and Table 1 have come to have first priority. Concepts such as artistic causality, idiosyncratic meaning, and intentional symbolization, as Beittel uses them to accent the artist as agent, to focus on the subjective meanings possessed by the artist, and to see the attempt to transform the meanings into any one of several concrete media, seem very powerful. They seem linked closely to an important and defensible definition of more general educational goals.

I feel a need to debate formally the issue more fully with those whose priorities are different, but who have been major contributors to CEMREL's curriculum, both products and theories. I also feel a need to present to the reader discussions of the "artists" series of packages—*Composer, Visual Artist, Storyteller, Architect,* and so forth, which have been produced since our observations of the program in action. But there is not space here. Similarly, the possible flow of the other experiences/roles/behaviors as antecedents to creativity has not been explored theoretically or empirically in the program. Nonetheless, the emphasis on creativity is an initial personal stand on the dilemmas in priorities. It begins for me the sequence of decisions leading to the judgment of effective teaching in aesthetic education.

A conception of the domain and improvising in the classroom. Labels are curious phenomena. At best, they represent major concepts; at worst, they are empty verbalisms or unrelatable nonsense syllables. Usually, they fall somewhere in between. For the initiated, acronyms are quick and easy means of communicating; for the uninitiated, they can be the worst of labeling. In some of the schools in our study, the aesthetic education program was known as "CEMREL." Teachers talked of the CEMREL program: "We have CEMREL at 11:00." Nowhere did we hear the term "aesthetics" being used. Usually, the alternative reference was to "Doing Dramatic Plot," the "Meter Box," or "Sound and Movement"; the statements were at the level of the concrete materials.

The pervasiveness and depth of this issue arose in one late evening recording of the Summary Observations and Interpretations notes:

> Another item that hit me, which may or may not be significant, and which may or may not be my own problem, is that no one in talking about the program today, really talked very clearly or abstractly or even to the point of the nature of aesthetic education. The people seem to be very much package-bound and not able to get beyond that in any fundamental way. I don't know whether it is my informal questioning or the kind of comments that the teachers make or what, but it always comes out as lan-

> guage development, language arts, reading, creative writing, etc. That one needs some careful cross-checking with Sally and her notes and also with the program people. I guess the feeling I have personally is that both in the formal documents, and in the discussions of the program people, there just isn't any real clear conception that overrides the totality. If that's true, then it might well be that some one conception ought to be the basic adoption for the run-of-the-mill making the points clear. Then the multiple alternatives be given in kind of an advanced AEP teacher training program. That may be all pie in the sky, too.

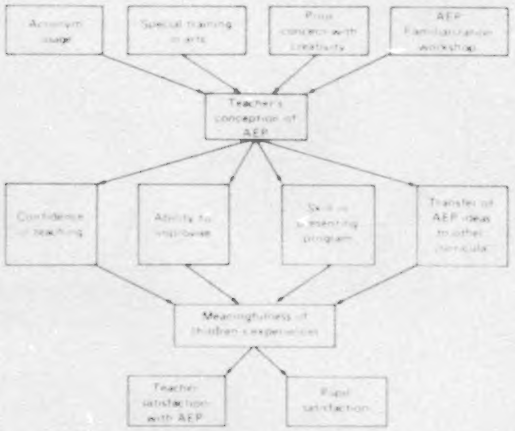
Some teachers we observed using the curriculum materials did something which left us with the impression, "They've got the concept." Usually, the teachers had a homey, almost slang, expression to accompany their directions to the children. One third grade teacher in beginning a *Dramatic Plot* activity, for instance, urged the children to "Make your story hang together." Another teacher using *Word Pictures* kept speaking of the "ideas" the word picture should convey and had the children draw their word pictures and note the differences when an adjective was changed. In Figure 3 we sketch some implications of teacher understanding of AEP.

Much of the psychological literature on concept attainment does the educator an injustice because it often deals with very simple and rudimentary concepts. A conception as broad and differentiated as aesthetic education and with elements relatable to so many facets of an individual's personal and professional life is neither easy to teach nor easy to attain. Further, the translation and transfer into overt teaching behaviors seems a very sophisticated process about which we know very little.

Scattered through Smith and Schumacher (1972) is a strong argument that AEP needed a model and accompanying language which the teacher could use in thinking about the specific elements of AEP as well as the totality. As we analyzed the field notes, an extension of that idea arose—the need for a language to communicate to the children. One of the indices we found ourselves using to determine if the teacher had a concept of the program was her utilization of quite concrete, often figurative or metaphorical expressions to the children. The children seemed to "get the idea" better when teachers improvised and talked that way about what the children should be doing.

FIGURE 3

Hypotheses Related to the Teacher's Conception of Aesthetic Education


Acronym usage
Special training in arts
Prior concern with creativity
AEP Familiarization workshop
Teacher's conception of AEP
Confidence of teaching
Ability to improvise
Skill in presenting program
Transfer of AEP ideas to other curricula
Meaningfulness of children's experiences
Teacher satisfaction with AEP
Pupil satisfaction

Several illustrations are scattered through this essay. A few are collected here to make the point explicit. One of the first times we witnessed the phenomenon was early in the fall with a teacher working with *Dramatic Plot*. We commented:

> In talking about this, it just occurred to me that one of the other differences in her class is that she talked much more explicitly about what the kids were creating as "a play." The other teachers often described it as a story. In this sense they were merging it with story writing, story telling, and creative writing of that sort. This teacher almost always kept coming back to the fact that it was a play and that had implications for what you would see if it were on the stage, or what you would have to do it it were on the stage, and so forth. Once again, to me, that's a very striking point of departure, a "set" to be given to the kids, that infuses the whole operation with a slightly different tone or perspective.

Another teacher in introducing *Dramatic Plot* gave a rapid fire series of directions: "Build up to the conflict on spaces 5, 6, or 7"; "Put meat on the skeleton, connecting the incidents..."; "Not acting out the story today...do a lot of talking so the people will know what is happening...."; "Get busy, make your story hang together." (Obs: She's got the concept.)

In retrospect, what we seemed to be saying extends our conception of creativity to what might be called creative teaching. The teachers we were resonating to were self-determining agents—teacher causality to transform Beittel's (1972) term.[9] They had a conception of the program, a kind of idiosyncratic meaning about aesthetic education as a curriculum. And through figurative, metaphorical, and "homey" expressions and improvising in their interactions with the children, they worked over those idiosyncratic meanings into concrete materials and actions, in the best sense of intentional symbolization. In short, they were creative teachers in a way consonant with being a creative artist.

Responsiveness to pupil ideas and perspectives. One of the long-term goals of education is the development of autonomous citizens. I would argue that teachers who are responsive to pupil ideas and perspectives tend to facilitate the autonomy of the child in a curricular domain (de Charms, et al., 1976). The observation which suggested aspects of this issue was conducted in a classroom in which a *Creating Word Pictures* lesson was under way.

> In with Mrs. Wald, a substitute for one of the teachers who is gone. She's already swinging with the program. Kids in 6 tables of 4. Room is bright and cheery. The teacher's very caught up. She tells me she's not "sure of the 'right' procedure"; I think she has the idea of clear, vivid, imaginative images. She's enthused, charged up.
>
> P: "I'm finished."
>
> T: "Now what are you going to do?"
>
> She has a confidence, etc. Kids are busy, chattering re the materials. Mrs. Wald says to me, "They are comparing, checking, and helping." She shushes them occasionally. The children have folded pieces of drawing paper into four squares (and draw each image): pink lovable sun, pink laughable sun.
>
> Mrs. Wald has the concept, e.g., when shushing the children she asks, "Terry, do you get all of your 'ideas' from your neighbor?" Accents ideas—"Laughing over silly ideas," etc. (Obs: (1) Capturing the intent of the idea is a major achievement, the use of pictures is very helpful; (2) recheck the emphasis on ideas in the teacher's guide; (3) somehow many of the other teachers seem to have been behaving more by rote through the workbook.)
>
> To one of the children she says, "Taste, touch, smell; all your senses get in there." She bounces around from one child to another. (Obs: Almost as though she's full of ideas and can extend any kid's idea.)
>
> Another child's work was observed and noted: "light fancy happy grass," "light plain lovely grass," etc. (Obs: This child has drawn grass with faces that are remarkably similar to *Characterization* materials. Linkage here could be made beautifully.)

We have chosen to accent a small part of a larger effort, the teacher's ability to perceive the child's intent and help him expand his ideas. The teacher seemed to have a clear conception of the program and combined her "disciplinary" interventions with this thrust, "Terry, do you...?" The children knew their verbal images were conveying ideas which they were also trying to represent in drawings. In all this, the teacher kept the class moving, both intellectually and managerially, by helping the children elaborate their products. Further complications arise because teaching and learning involves more than ideas—social and interpersonal skills, intellectual skill training and practice, development and expression of attitudes and feelings, and so on.

Nature and assessment of pupil learning. To this point, I have commented only implicitly about pupil learning. Since the area is a large one, I'll only try to outline some of the critical ideas. It always is tempting to speak of instrumentation as the evaluator's "thorniest problem." Like most curriculum laboratories, CEMREL has been faced with the strain between program and evaluation staff: "The program staff can't tell me their behavioral objectives; therefore, I can't measure what they want," and "The tests the evaluator creates miss the heart of the curriculum experience." In my judgment, these comments are in the domain of "validity" of measurement; a solution to the problem cannot even be approached without a clear conception, model, or construct of aesthetic education.

First, the models in Figure 1 and Table 2 seem appropriate as a specification table (Tyler, 1950) for the kind of pupil changes to be tapped. The pupil roles might legitimately be called "behaviors," as the behavioristic psychologist Berlyne (1971) explicitly does label them. The content areas have features unique to each of the arts. If the analogy holds, the evaluator has only to shift levels of specificity and concreteness. A creator role or a critic role has components that can be specified. This is no different, in our judgment, from saying that knowledge as a category is composed of facts, concepts, and principles and that intellectual skills as a category are composed of analysis, synthesis, and evaluation, as Bloom et al. (1956) do in the *Taxonomy*.

One moves as specifically or as abstractly as one's problems and purposes demand. For some purposes, we may well need to specify the elements of each role. For instance, what are the knowledges, skills, orientations, and so forth of the drama critic? What do these components mean at the third grade level or the sixth, ninth, and twelfth grade levels? How are they different from the components of the third or twelfth grade playwright. Test and measurement types, at least those who worry about achievement tests in school learning, base much of their argument on content validity of the measures. Content validity is an attempt to attest that one's measure samples adequately the domain of the course or curriculum.

A more powerful approach, construct validity, has been suggested by the APA Committee on Psychological Tests (1964). Essentially, this approach involves both a theoretical and empirical attack on measurement problems. A theoretical or nomological network is sketched out, experiments are designed to coordinate with the theoretical system, and results are obtained. The clarification of constructs, hypotheses, and operational indicators moves forward concurrently. A critical element in this is the need for theoretical models of the events involved.

We are arguing that our models attempt to state a theoretical structure of AEP. With that structure in mind, the problem becomes coordinating operational modes with the theoretical structures. To this point, since there has been no middle level theory which would permit such an analysis, it has not been possible to speak of the construct validity of any of the measures so far developed. The argument we have been trying out is that the models are a way to begin those discussions which will enable us to set some superordinate goals and resolve some of the existing conflict.

The model leaves unsolved two quite critical and interdependent problems—the level of learning or personality change sought by the program and the kind of theory into which the changes will be cast. Developmental psychologists such as Gardner (1973) speak of underlying structures, stages, modalities, and factors which are derived from such theorists as Piaget, Levi-Strauss, and Erickson. These seem very different from the child's learning the concepts of duple or triple meter or his ability to perceive these in a piece of music, to define character, setting, and incident in a dramatic plot, to create a plot with these elements, or to critique a classmate's efforts. The theoretical linkages between these kinds of learnings and their susceptibility to formal instruction are critical, difficult, and unsolved problems for an analysis of effective teaching.

The "kind of theory" seems almost another way of saying the same thing. In aesthetic education, there are frequent quarrels between evaluators who tend to take a more outside/behavioral viewpoint—what can the child now do?—and many of the curriculum developers and teachers who take a more internal/experiential view—what is happening to the child's point of view? The aesthetic world is full of items like expression, metaphor, intrinsic meaning where there is a manifest or overt statement and meaning as well as a latent or covert meaning. Such phenomena seem much more difficult to handle in a descriptive behavioristic language. Practically, the problems are even more acute with a relatively unsophisticated behavioral approach.

Converting these issues to empirical problems has left us with some data but mostly hunches. The overall assumption we have made is that operational definitions can be made of the concepts implied in the models. The instrumentation we have argued most strongly for involves three broad strands which we would hope to triangulate (Smith, 1974; 1975). They are: (1) Piagetian type clinical interviews; (2) videotape recording and content analysis of performance and process data; and (3) product analysis of artifacts produced by pupils. Each seems susceptible to both qualitative and quantitative analysis. Once again, I'm impressed with Beittel's beginnings in his concern for the artists' creative stream of consciousness, approached through a special participant observer role.

In short, particularistic stands on the valuation/theoretical/empirical issues can be taken and defended. Presumably they can be made operational, and judgments of effective teaching can be made in terms of them.

Some final thoughts. Over the years, the naturalistic qualitative inquiry stance has gotten us close to important practical and theoretical problems in urban education, in educational innovation and in curriculum evaluation. Most recently we have focused on aesthetic education. Through each problem and setting, we have explicitly dealt with teaching but only implicitly addressed "effective teaching." This essay has attempted to redress that focus.

The thesis that evolved in thinking through the problem and in shaping the structure of the argument has been that "effective teaching" is a complex valuational/theoretical/empirical judgment. At a minimum, this seems to involve: (1) a general conception of education and teaching; (2) a conception of a curriculum domain (e.g., aesthetic education); (3) a set of priorities in that domain and in relation to other domains (e.g., language arts, the total elementary curriculum); (4) attention to the possible conflict in values and priorities among relevant groups; (5) a realization that instrumentalities are intriguingly complicated by functional equivalence; (6) an eclectic language structure about teachers, classes, and children that nearly defies rigorous thought; (7) a multimethod, multiperson, multisituation, and multivariable approach to data collection, combination, and reporting; and (8) a particularistic but defensible stand.

Notes

1. Part of this work was supported directly by CEMREL, Inc., and indirectly by NIE, USOE, and the Fulbright-Hayes Research Fellowship Program. The opinions expressed here do not necessarily reflect the positions or policy of any of these organizations; no official endorsement should be inferred.
2. The *Meter* package is a set of lessons introducing children to duple and triple meter in music. It is one of some forty projected packages, each involving ten to fifteen hours of multimedia instruction in multiple experiences (creating, performing, implementing, appreciating, and critiquing) across multiple art forms (art, music, drama, dance, literature, photography) and into cultural and environmental applications.
3. Early on, we phrased this as teacher decision making (Smith and Geoffrey, 1968). More recently, Harold Berlak's writing and helpful conversations have been especially provocative in extending the theoretical ideas (Berlak, 1963; Shaver and Berlak, 1968).
4. The *Creating Dramatic Plot* package involves children in small cooperative groups working in a graded series of game-like activities wherein they construct dramatic plots containing such elements as characters, settings, incidents, conflicts, crises, and resolutions.
5. *Creating Word Pictures* is a package which strives through game-like activities to teach the children to develop imaginative and novel word images.
6. The role of improvisation in aesthetic education is currently a major topic of contention and discussion (Cheifetz, 1971, and Sutton-Smith, 1971).
7. In a sense I am raising the aptitude/treatment interaction problem at the level of teachers and materials, in contrast to the usual pupil and materials interaction.
8. A variety of sub-arguments and data are raised here—e.g.,

the argument on the relative influence of motivational versus intellectual factors in creative achievement.

9. Much of Beittel's discussion is based on de Charms (1968, 1976) provocative work.

References

American Psychological Association, Committee on Psychological Tests. *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (rev. ed.). Washington: The Association, 1964.

Ausubel, D. *The Psychology of Meaningful Verbal Learning.* New York: Grune and Stratton 1963.

Barkan, M., et al. *Guidelines: Curriculum Development for Aesthetic Education.* St. Ann, (Mo.): CEMREL, 1970.

Becker, H.S "Problems of Inference and Proof in Participant Observation." *American Sociological Review* 28:652-60, 1958.

Beittel, K.R. *Mind and Context in the Art of Drawing.* New York: Holt, Rinehart and Winston, 1972.

Beittel, K.R. *Alternatives for Art Education Research.* Dubuque: Wm. C. Brown, 1973.

Berlak, H. "The Construct Validity of a Content Analysis System for the Evaluation of Critical Thinking in Political Controversy." Unpublished dissertation, Harvard University, 1963.

Berlyne, D.A. *Aesthetics and Psychobiology.* New York: Appleton-Century-Crofts, 1971.

Bloom, B.S., et al. *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain.* New York: Longmans, Green, 1956.

Broudy, H.S. *Enlightened Cherishing: An Essay on Aesthetic Education.* Urbana: University of Illinois Press, 1972.

Bruner, J. *The Process of Education.* Cambridge: Harvard University Press, 1960.

Campbell, D.T., and D.W. Fiske. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56:81-105, 1959.

Charters, W.W. Jr., and J.E. Jones. "On the Risk of Appraising Non Events in Program Evaluation." *Educational Review* 2:5-7, 1973.

de Charms, R. *Personal Causation.* New York: Academic Press, 1968.

de Charms, R., et al. *Enhancing Motivation Change in the Classroom.* New York: Irvington, 1976.

Chiefetz, D. *Theater in My Head.* Boston: Little, Brown, 1971.

Denzin, N. *The Research Act: A Theoretical Introduction to Sociological Methods.* Chicago: Aldine, 1970.

Dunkin, M., and B. Biddle. *The Study of Teaching.* New York: Holt, Rinehart and Winston, 1974.

Gardner, H. *The Arts and Human Development.* New York: Wiley, 1973.

Green, T.F. "A Topology of the Teaching Concept." In R.T. Hyman (ed.). *Contemporary Thought on Teaching.* Englewood Cliffs (N.J.): Prentice-Hall, 1971.

Hirst, P.H. *Knowledge and the Curriculum.* London: Routledge and Kegan Paul, 1974.

Kounin, J.S. *Discipline and Group Management in Classrooms.* New York: Holt, Rinehart and Winston, 1970.

Madeja, S. "Aesthetic Education: An Area of Study." *Art Education* 24(8):16-19, 1971.

March, J., and H. Simon. *Organizations.* New York: Wiley, 1958.

McLean, L. "Judging the Quality of a School as a Place Where the Arts Might Thrive." In R. Stake (ed.). *Evaluating the Arts in Education: A Responsive Approach.* Columbus: Charles Merrill, 1975.

Meehl, P.E. *Clinical vs. Statistical Prediction.* Minneapolis: University of Minnesota Press, 1954.

Merton, R.K. *Social Theory and Social Structure* (rev. ed.). Glencoe (Ill): Free Press, 1957.

Peters, R.S. "Education as Initiation." In R.D. Archambault (ed.). *Philosophical Analysis and Education.* London: Routledge and Kegan Paul, 1965.

Richardson, E.S. *In the Early World: Discovering Art Through Crafts.* New York: Random House, 1964.

Rosenshine, B. *Teaching Behaviors and Student Achievement.* London: National Foundation for Education Research in England and Wales, 1971.

Scheffler, I. "Philosophical Models of Teaching." In R.T. Hyman (ed.). *Contemporary Thought on Teaching.* Englewood Cliffs (N.J.): Prentice-Hall, 1971.

Shaver, J.P., and H. Berlak (eds.). *Democracy, Pluralism, and the Social Studies.* Boston: Houghton Mifflin, 1968.

Smith, L.M. "Images and Reactions: A Personal Report on 'An 8 Day Week.' "In S. Madeja, et al. (eds.). *Working Papers: Final Report on the Institute in Aesthetic Education for Administrators.* St. Louis: CEMREL, 1974a.

Smith, L.M. *The Child's Aesthetic Experience: An Interim Report and Position Paper.* St. Louis: CEMREL, 1974b.

Smith, L.M. "Psychological Aspects of Aesthetic Education: Some Initial Observations." *Bulletin* 43:92-115, Council for Research in Music Education, 1975.

Smith, L.M., and J.A.M. Brock. "Go, Bug, Go!": *Methodological Issues in Classroom Observational Research.* Occasional Paper #5. St. Ann, (Mo.): CEMREL, 1970.

Smith, L.M., and W. Geoffrey. *The Complexities of an Urban Classroom.* New York: Holt, Rinehart, and Winston, 1968.

Smith, L.M., and J. Greenberg. *Analytical and Empirical Schemes In the Arts and Aesthetic Education: The Collision of Theory and Data.* In process.

Smith, L.M., and P. Keith. *Anatomy of Educational Innovation.* New York: Wiley, 1971.

Smith, L.M., and P.A. Pohland. "Education, Technology, and the Rural Highlands," In D. Sjogren (ed.). *Four Evaluation Examples: Anthropological, Economic, Narrative, and Portrayal.* Chicago: Rand McNally, 1974.

Smith, L.M., and S. Schumacher. *Extended Pilot Trials of the Aesthetic Education Program: A Qualitative Description, Analysis, and Evaluation.* St. Louis: CEMREL, 1972.

Solomon, W. "Toward the Development of a Theory of Teaching: A Participant Observer's Study of the Experimental Trial of Lessons from a Social Studies Curriculum Project." Unpublished dissertation, Washington University, 1971.

Sutton-Smith, B. "A Developmental Psychology of Play and the Arts." *Perspectives on Education* 4(3):8-17, 1971.

Toffler, A. "The Art of Measuring the Arts." *Journal of Aesthetic Education* 4:53-72, 1974.

Tyler, R.W. *Basic Principles of Curriculum and Instruction.* Chicago: University of Chicago Press, 1950.

Zetterberg, H. *On Theory and Verification in Sociology* (3rd ed.) Towtowa (N.J.): Bedminster Press, 1965.

CRITIQUE

Gary A. Griffin
Teachers College
Columbia University

My critique of the papers by Cooley and Smith is guided by certain concerns about our efforts to understand the complex phenomenon of schooling.

First, I am concerned that although Meehl's comments, quoted by Cooley, were written more than two decades ago, they still characterize the education community. The polarization of those who depend upon quantitative research methods and those who rely on other methods, variously named but called qualitative for this conference, is alive and well. Indeed, the polarization is evident throughout the papers presented here. But it is not limited to research methodology. Witness the competency/humanist controversy in teacher education, the head-on collision of the psychologists and neo-romantics about the nature and consequences of school life, the contrast between "teacher-proof" materials of instruction and those presented as teacher resources. These disparate examples illustrate what I consider to be a persistent dilemma: the dependence upon and belief in one way of knowing over another; the rejection, on the one hand, of what is known as a consequence of measuring, quantifing, reducing, and numbering and, on the other hand, of prose descriptions, the logic of oral and written language as revealing of what is being studied.

Second, I am concerned that the user of research must shoulder the synthesizing chore. To reduce via number alone is nonsensical, and to capture the whole with endless description when methods exist to realistically and efficiently encapsule is to ignore some of the best tools we have.

Third, I am concerned that I cannot observe in schools across the country any widely observable consequences of the findings of teacher effectiveness research conducted in laboratory settings. I cannot find to any great extent that desired phenomenological "match" between what is controlled and controlled for in laboratory settings and what is observed in schools—whether in midtown Manhattan or Anniston, Alabama.

And fourth, I cannot in conscience rationalize the ideological and pedagogical distance between the researcher-theoretician and the teacher in a school by blaming the practioner for not "keeping up." Keeping up with what? Is it a reasonable proposition that the researcher should accomodate his language and communication system to the client? I think so. And is it reasonable that our research should be relevant to those who are expected to use it? Again, I think it is.

It is my stance that research designed to identify effective teaching should be useful to researchers and other disciplined inquirers in understanding what teaching is and in acting upon what is. It should reflect what has come before and line up with or perhaps push us into what can be. It should promote a recognition of what is believed to be so; if you will, cause us to say in Phil Jackson's language, "Yes, that's life in classrooms." With this in mind, let me turn more specifically to the two papers.

Cooley's call for a careful identification of the requirements of research on effective teaching as the elemental issue which informs our selection of method is, to me, the best of beginnings. His attention to what we have depended upon as outcome measures links us methodologically and substantively to our research history, calling to question the limitations of that history and pushing us outward from it. The paper serves as a good sorting device, guiding us through some of our longstanding bugaboos and pitfalls. It links us to our quantitative research past, presents some powerful lessons learned over the years, and moves us to the consideration of method which takes both into account, not as roadblocks but as road signs.

Professor Smith's paper illustrates through example the productivity of qualitative method. His rich vignettes reveal and raise questions; they cause the inquiring reader to speculate and want to test. In combining description of and reflection upon his observations, the narrative helps us to clarify the process of analysis, synthesis, and evaluation so central to the tasks of the educational researcher. It is this replication of what is done as well as what is seen that provides the seeds of a new communication system between researcher and client. Even though the concerns and assumptions of the researcher may differ from those of the researched or subsequent users of the findings, the context of the presentation—written language—does not require the learning of a new technology.

The Smith paper also helps us to acknowledge the power of a method which pulls together, rather than pulls apart, the qualitative and quantitative. His discussion of the construction of the multimethod, multiperson, multisituation, multivariable matrix of data is a particularly fine, explicit example of this.

Both papers call attention to the difficult task of gaining access to schools, intervening into the lives of those in them, and gaining an authentic representation of what occurs. I think we must listen carefully to teachers and other "out there" who tell us that what we do is irrelevant; that our interests are esoteric; that our methods and procedures are arcane; that our messages—couched as lessons for them to learn—are not understandable. We must come to agreement that we will organize and act *with*. Whether our methods be characterized as quantitative or qualitative, we must gain credible access to the system we study. And that credibility will, I think, come as a consequence of mutual deliberation and decision between those who study and those who are being studied. If we work toward discovery with teachers, we also can work toward change with teachers.

A few words about the term "effective teaching." As Cooley points out, the effect most often is seen to be upon student learning as measured by some valid, reliable, ethically constructed and administered set of instruments. Along with Cooley and Smith, I am concerned about the rather narrow conception of teaching this definition represents. As both papers describe, teachers plan, provide materials, interpret events of classroom life, monitor and report on student behavior, group, create social climate, set rules, and so forth. Perhaps by broadening our concept of teaching to include those acts of teaching which are not easily described as instructional talk, we could discover a *set of conditions* (no matter how brought about by teachers) which relate to pupil outcomes.

Teachers do so much else besides talk at or to or with students. We can at least begin to sort out what effect the teacher has on that "much else." It is here that I applaud the wisdom of Professor Smith in his deceptively simple declaration that effective teaching is a complex "theoretical/valuational/empirical judgment." His positive feelings about the potential of such a stance is encouraging, indeed.

I share this positive attitude. My hopefulness is strengthened by the careful attempt of both the Cooley and Smith papers to sort out and comment upon our past, point to our present dilemmas, and suggest means for acting on those dilemmas. And, importantly, both papers call upon us to withdraw from our either/or, qualitative/quantitative, off/on, yes/no positions and move toward a mode of inquiry which has as a major component the selection of procedures based in appropriateness, whether the consequence be qualitative, quantitative, or a juxtaposition of both.

ASSESSING RACE RELATIONS IN THE CLASSROOM

Court-ordered busing of students in order to achieve equal educational opportunity has dramatically altered the classrooms of inner-urban America. Forced together and expected to learn are disparate groups of children, each with differing cultural backgrounds and expectations for schooling. Within such a mixed-group context, what is "disadvantaged" for one culture may not be "disadvantaged" for another. Understanding the effects of schooling under such circumstances can be enhanced by understanding intergroup relations, in particular between children and teachers of different races. How can we best assess race relations in the classroom and their effect on schooling?

RACIAL TENSION IN HIGH SCHOOLS
PUSHING THE SURVEY METHOD CLOSER TO REALITY

Robert L. Crain
The Rand Corporation

Educational research has always contained a conflict between the proponents of quantitative and qualitative methodologies. The distinction, as used in this paper, has to do with *when* the decision is made about what variables should be studied and how they should be measured. In quantitative research, the variables must be selected in advance, and most details of the measurement technique must be known before research begins. Granted, the quantitative researcher normally includes a wide variety of variables in hopes that a large net will capture the interesting processes, but completion of the questionnaire locks out the possibility of adding new variables. The qualitative researcher, on the other hand, is free to go into the field with a very loose set of notions in hopes that observation will help him discover the critical variables.

If one begins with these definitions and asks, "Which method is preferable?" the answer is obvious—each has strengths and limitations, and different problems require different approaches or a different mix of the two approaches. While this sounds more like a recipe for division of labor than for conflict, every social science discipline with both "hard" and "soft" researchers is characterized by a sometimes bitter controversy. Why does a choice of methodology generate such conflict? Partly, it is debate for its own sake. But in addition, the choice of method generates conflict because it influences the kind of research which can be done and even affects the ideological predisposition of those who perform the research.

Table 1 is an effort to understand the intellectual and ideological baggage which seems to accompany the choice of research method. It defines quantitative and qualitative researchers as Weberian (1930) ideal types, and no real person fits an ideal type. Nonetheless, it is a useful heuristic tool for understanding what the limitations of each method do to research and researchers. For example, the quantitative researcher is limited to those variables already identified in the literature and those which are accessible. There are many measures of student socioeconomic status in surveys, but many fewer measures of teacher practices; such variables are expensive and require the permission of teachers. Quantitative research also tends to focus upon the standardized achievement test for the same reason—the tests are routinely administered, and often are available at little or no cost.

The ramifications of the choice of measurement are widespread. Quantitative data permit elaborate statistical analyses. At the same time, the demand of statistical rigor may influence the analyst to avoid going "beyond the data" into "speculation." But what is speculation for a quantitative researcher may be theoretical argument for a qualitative researcher. Since the qualitative method requires the presentation of case material and does not permit elaborate statistical analysis, the researcher must necessarily find something to say, and the something is verbal, not quantitative. Consequently, the ideal type qualitative analysis is a mixture of case material and theoretical argument. The fact that qualitative research has more visible theory reflects more the demands of the research than the greater power of the method to produce theoretical conclusions.

Quantitative methodologies lend themselves to correlational analysis—the study of the relationships between variables. The statistical techniques permit dealing with large numbers of variables simultaneously in multivariate analysis. On the other hand, the qualitative researcher can describe quite accurately the measurement of a particular variable. The quantitative researcher may use an elaborate index to measure racial interaction, but scores on the index are likely to have no intuitive meaning. The qualitative researcher can present a series of incidents from the observed

TABLE 1

Contrasts of Quantitative and Qualitative
Educational Researchers as
Weberian Ideal Types (or as Stereotypes

	Researcher Type	
	Quantitative	Qualitative
Variable selection	Limited to known & accessible variables	Add variables in the field
Cost	High	Low
Sample size	Large	Small
Type of control of error	Reliability	Validity
Analysis approach	Statistics or logical modeling (e.g., Boolean algebra)	Theoretical argument & verbal presentation of incidents
Principle analysis method	Correlational (causal)	Measurement (descriptive)
No. & source of variables in analysis	Stresses multivariate relationships among old variables	Defines new variables, finds two-variable relationships
Interaction effects	Few	More
Theoretical perspective–discipline	Psychometrics, economics, psychology, sociology	Political science, anthropology, sociology
Theoretical perspective–concepts from sociology	Learning theory, attitudes, survey research, social stratification, organization theory, experimental social psychology	Socialization, functionalism, symbolic interaction, culture, norms, ethnomethodology
Data writeup	Tables & Interpretation	Theory & case material
Stress on quality of writing	Little	More
Ideological perspective	"Value free": incrementalism, conservative, scientific	Value laden: global reforms, radical, humanistic

situation, detailing them so as to make "real" to the reader precisely how much interaction occurs. Thus, the qualitative research report is likely to emphasize the theoretical definition of a new concept and its measurement, perhaps with a lengthy discussion to point out that the "scores" on this new "variable" are higher than one might expect and a brief discussion of how this variable is linked to two or three others. The ideal type quantitative analysis mainly uses existing variables, pays little attention to the absolute magnitude of the scores, and stresses complex multivariate causal modeling.

The quantitative researcher must self-consciously intend to find interaction effects wherein for a portion of the sample a causal relationship is of one kind and for another, different. The qualitative researcher finds it much easier to recognize that a relationship holds in one case but not in another (presuming he has more than one or two cases to study). The qualitative researcher, studying a single case, tends toward arguing that the case under study is typical and the description there fits everywhere—the individual differences are less interesting than the continuity.

Researchers who use qualitative and quantitative methods receive different training, read somewhat different materials, and consequently draw upon different theoretical disciplines. Quantitative researchers in education are likely to be trained in psychology, psychometrics, or economics; they rarely have backgrounds in anthropology or political science. When sociological tools are borrowed for educational research, the qualitative researcher has a somewhat broader range of theoretical argument. He has, for example, the work of Howard Becker (1961) and others on socialization and role theory. He also can use the classical sociological literature and draw on the way ethnomethodologists use the philosophy of meaning. The quantitative educational sociologist is likely to be a survey researcher, familiar with research on attitudes, the relationship between attitudes and behavior, and social stratification. He may be able to apply some work from organization theory. He is likely to be more familiar with research in experimental social psychology and be quick to pick up the work of Rosenthal (1968) to test in a survey of schools. There is no logical necessity that particular theories require particular methods—a point Stinchcombe (1964) makes in a humorous paper pointing to a very large number of quantitative studies which might test the functional theory of stratification. Stinchcombe implies that the failure of empiricists to use this particular theoretical approach reflects their shortsightedness much more than limitations inherent in the empirical method.

The type of method used influences the presentation of conclusions which in turn influences the audience to which the writer must speak. Quantitative research necessarily implies tables and statistics, and it is the rare quantitative researcher who can make his research comprehensible to a wide audience.

But perhaps the most important difference is in ideological perspective. Quantitative work uses its rigor as an argument for objectivity in science, and the quantitative researcher normally takes great pride in controlling his own biases. The qualitative researcher has little protection against the influence of his own bias and turns the argument about scientific objectivity on its head in defense, arguing that value free research is biased toward establishment values. This charge is partly true, because the quantitative researcher must necessarily compare what exists in one place to what exists in another—he cannot compare what exists to what should exist in a better world. Consequently, his approach to educational reform must be incrementalist.

Conversely, it is not accidental that the writers on educational reform whom Havinghurst has characterized as "educational anarchists" almost without exception disdain the quantitative approach. If the system is bad at its roots, comparing one leaf to another will not get us very far.

The point of Table 1 seems to be that a whole variety of social processes have gone into creating these two types of researchers. It may have begun in graduate school when students discovered that some were good at statistics and others good at theory and arbitrarily dichotomized the world into people who could do one and not the other. This method stereotyping is not unlike the sex stereotyping of pre-adolescent children. Quantitative researchers learn to say, "I can't do theory," in much the same way boys learn to say, "I can't cook."

While this table was interesting to construct, it should not be taken very seriously. There is a rich variety of research within each methodology, and the ideal types described here are figments of my imagination rather than empirical descriptions. While a commitment to quantitative research exerts pressure to use standardized achievement tests which are widely available and have known reliabilities, more than a handful of quantitative researchers have rejected standardized testing. Likewise, there are qualitative researchers whose views on educational innovation are very much in the incrementalist tradition. Most important, I suspect that those researchers who are less easily placed into these neat categories are the most valuable. Gerald Suttles (1968), perhaps the best fieldworker in sociology, has an advanced degree in mathematics, and the principal investigator of one outstanding quantitative research project has a deep philosophical commitment to statistics but very little skill in carrying out the actual addition and subtraction.

The danger with Table 1 is that it reinforces our natural tendencies to stereotype the two sides of the agrument. This paper seeks to demonstrate that someone who describes himself as a "quantitative" researcher is not necessarily guilty of all the crimes in the stereotype of Table 1. The data analysis which follows is presented not for its theoretical value but to demonstrate what quantitative work can do. The analysis is sensitive to the methodological issues that characterizes all good quantitative research; it is concerned with validity as well as reliability, sensitive to the possibility of interaction effects within the universe, and so forth. But it also tries self-consciously to break out of the style of research which comes naturally to quantitative work. It ignores standardized achievement tests in favor of a noncognitive measure and searches for its theoretical explanations in some of sociology's nonquantitative traditions.

Probably the most important characteristic of this paper is that there is very little in it which the researcher anticipated at the beginning of the study. The quantitative method does, indeed, free the researcher from his own biases. No matter how strongly the research instrument may have been biased to produce a positive relationship between two variables, there is still the possibility that the relationship will be negative.

But if the paper demonstrates the strengths of the quantitative method it also demonstrates its weaknesses. The intense amount of statistical analysis behind this paper diverts energy from working toward a theoretical understanding. Consequently, this paper seems to tell us a lot we did not know about racial tension in southern high schools but does not satisfy our need for even the beginnings of a coherent theory of race relations in schools.

The Study

The data presented here are taken from a survey of southern high schools by the National Opinion Research Center (1973), conducted as part of a major experimental evaluation of the Emergency School Assistance Program. The study is important because it represents the first use of randomized experimentation in evaluating a large federal education program (see Crain and York, 1975, for a description of the experiment). The data were gathered from a survey not of individuals but of schools. In each of 200 southern high schools, the principal, 10 teachers and more than 50 white and black students were given questionnaires, and their combined response was used to describe each school.

Underlying the research instrument is the concept of the school as a social organization with a social climate and the belief that a portion (though not all) of this social climate is affected by the conscious and unconscious actions of the superintendent, principal, and teachers. Thus, this research is in the intellectual tradition of Coleman (1959) and McDill (1965).

Lengthy questionnaires measured a number of racial variables in several ways. We asked teachers about their attitudes regarding race issues in general and school integration specifically. We asked everyone—principals, teachers, and students—to report on teachers' behavior regarding race relations. We also asked students to describe their own racial attitudes and level of racial contact with other students. This makes it possible to measure for each school a variety of components of the school racial climate: the average attitudes and racial behavior of all teachers and the racial attitudes and behavior of the student body as a whole. Our method hinged upon recognizing that each subject was both a respondent, answering questions about his own attitudes, and an informant, with inside knowledge of the school. We were careful to distinguish between attitudes and behavior; we wanted to study the actual behavior of actors in the school and thus asked a number of questions about behavior and especially asked our informants to describe the behavior of others. The principal and teacher questionnaires also told us about the school's racial history, its use of tracking, and even the win/loss record of its athletic teams. The student questionnaire included questions on when the students were first desegregated, whether they were being bused past the nearest school, measures of socioeconomic status, and a short test measuring knowledge of black history.

We tried to be sensitive throughout to the biases different types of respondents would have. A good rule in survey analysis is never to take a questionnaire response at face value. The survey questionnaire should be seen as a series of micro experiments. The subject is given a stimulus, in this case a written question, to which he reacts by choosing one of several answers. It is then the job of the analyst to decide what it means when a certain fraction of subjects gave a particular answer. In some cases even careful wording could not overcome the potential for biased responses, and we turned for information to people who were not connected with the school at all. The interviewer was asked to double as an observer and report on various aspects of the school, including its physical condition. Telephone interviews were also conducted with a biracial panel of community leaders who were asked about the community reactions to desegregation.

TABLE 2

Means, Ranges, Standard Deviations, and Weights of Tension Items

Variable Description	Mean	Range	σ	Weight
1. % W saying there are few or no problems between blacks and whites in the school	78.2	0-100	17.4	−.174
2. % W saying tensions have made it hard for all	50.0	0-100	22.8	.050
3. % W reporting black complaints of favoritism toward whites	57.8	0-100	23.0	.058
4. % W reporting white complaints of favoritism toward blacks	50.6	0-100	21.9	.056
5. % W reporting white student attacks on black	13.2	0-100	14.5	.068
6. % W reporting black student attacks on white	40.0	0-100	30.9	.100
7. % B saying there are few or no problems between blacks and whites in the school	81.2	0-100	17.8	−.091
8. % B saying tensions have made it hard for all	47.7	0-100	19.6	.087
9. % B reporting black complaints of favoritism toward whites	60.8	0-100	23.6	.080
10. % B reporting white student attacks on blacks	24.2	0-100	23.4	.066
11. % B reporting black student attacks on whites	31.2	0-100	27.1	.152
12. % T saying desegregation has created no problems or some minor problems	81.4	0-100	18.1	−.325
13. % T reporting more fighting than before desegregation	26.1	0-100	24.4	.165
14. P: count of number of students treated in hospital	.16	0-4	.59	1.09
15. P: count of number of students treated by MD or nurse	.71	0-4	1.20	1.09
16. P: count of number of locker break-ins	2.35	0-4	1.99	1.09
17. P: count of number of gang robberies of students	.52	0-4	1.21	1.09
18. P: count of assaults on teachers by students	.13	0-4	.53	1.09
19. P: count of robberies of school property of over $ 0	.89	0-4	1.26	1.09
20. P: was school closed because of disturbances? (1 = yes, 0 = no)	.03	0-1	.28	13.6

The Measurement of Racial Tension

All informants were asked in detail about the level of racial conflict in the school and a racial tension was constructed. Table 2 lists the items in the scale, the approximate wording, and the mean response for all 175 schools. For example, students were asked, "On the whole, how would you say things are working out with both blacks and whites in the school?" and given four alternatives. The first line of Table 2 shows that 78 percent of the white students said "almost no problems" or "some minor problems"; 22 percent said "some serious problems" or "many serious problems." Lines 7 and 12 give the views on this question of blacks and teachers; 81 percent of both groups said "few" or "almost no problems." While this suggests the average school in the sample does not have serious problems, some other answers are less encouraging.

About half of the students said both that tension has made going to school difficult and that there has been a racial protest. About a third of the black and slightly more white students reported attacks on whites by blacks; about a quarter of blacks and a smaller number of whites reported whites attacking blacks. Thus, while each group tends to be somewhat biased, blacks as well as whites reported more black assaults. Only a quarter of the teachers reported increased fighting since desegregation. In the typical school, the principal could report only a single incident of a student being injured in a fight, although in one out of five cases, the student was sent to a hospital for treatment. There were occasional cases of gang robberies and rare attacks on teachers. Finally, three percent of the schools were closed because of racial tension. On balance, it seems clear that the typical southern desegregated school does not have a large amount of racial difficulty. Students

TABLE 3

Correlations Among Tension Measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. W: Few problems*	--	.74	.53	.49	.44	.71	.32	.31	.39	.42	.64	.53	.48	.24	.30	.01	.28	.22	.33	.03
2. W: Tensions		--	.75	.77	.54	**.76**	**.29**	.34	.41	.50	.58	.48	.49	.19	.29	.02	.25	.15	.24	.21
3. W: Black complaints			--	.72	.40	.63	.28	.27	.51	.34	.45	.39	.41	.11	.31	.06	.20	.14	.22	.30
4. W: White complaints				--	.43	.59	.18	.25	.28	.39	.35	.32	.32	.06	.12	.04	.10	.01	.16	.22
5. W: White attacks					--	.70	.31	.29	.34	.72	.64	.31	.45	.16	.33	.08	.27	.24	.25	.22
6. W: Black attacks						--	.31	.36	.42	.56	.80	.48	**.56**	.24	.36	.10	.32	.23	.31	.23
7. B: Few problems*							--	.30	.35	.29	.28	.30	.30	.17	.11	.06	.09	.06	.04	.15
8. B: Tensions								--	.40	.34	.43	.29	.38	.08	.17	.07	.18	.17	.11	.21
9. B: Black complaints									--	.28	.43	.28	.34	.12	.27	.10	.16	.14	.17	.27
10. B: White attacks										--	.67	.27	.41	.10	.08	.03	.12	.15	.05	.12
11. B: Black attacks											--	.47	.59	.19	.32	.12	.34	.25	.25	.28
12. T: Problems are minor*												--	.57	.27	.27	.14	.31	.33	.25	.16
13. T: More fighting													--	.13	.19	.16	.20	.24	.17	.18
14. P: Hospital treatments														--	.41	.17	.31	.15	.16	.06
15. P: MD/Nurse treatments															--	.13	.38	.21	.34	.28
16. P: Locker robberies																--	.17	.00	.09	.09
17. P: Gang robberies																	--	.31	.17	.19
18. P: Assaults on teachers																		--	.27	.03
19. P: Property robberies																			--	.22
20. P: School closed																				--

*Variable reversed

are uncomfortable to some degree, but they are hardly in any danger.

The twenty items of Table 2 are positively intercorrelated as shown in Table 3. With the exception of only one item (the principals' reports of robberies from lockers which was retained in the scale by error), all the correlations are positive. Many are quite large; for example, the correlation between the percentages of white and black students reporting attacks on whites by blacks in their school is .80. Teacher reports of increases in fighting were correlated around .5 with black and white reports of violence. Combined, the items produce a scale with a reliability coefficient in excess of .9, unusually strong by survey standards. One might question the inclusion of several student reports of protest activity. Presumably, protests about mistreatment are merely the exercise of democratic rights and should not carry a negative connotation. However, the fact is that reports of protest activity are positively correlated with reports of violence by students, teachers, and principals. The protest items are as much a part of the scale as the violence items.

The scale was built by multiplying each item by the weight shown in the right-hand column of Table 2, so that each type of respondent (black students, white students, teachers and principals) contributed equally to the scale.

The tension measure, then, has face validity and high reliability. It represents the views of different types of informants; it combines perceptions of action that has occurred with feelings about the tone of the school; the measures from the different informants are highly correlated. But the measure will be useful only if it helps us learn more about racial problems in schools. This is not to say that we should expect these data to produce a simple and coherent explanation for racial tension or a "quick fix" for school problems. There are many theories of racial conflict and part of the problem with any analysis of tension is that they all are true to some extent.

The relationship between racial tension and school racial composition is not linear. In Table 4 the greatest amount of racial tension occurs in racially balanced schools, while predominantly white and predominantly black schools have lower tension. Moreover, these three types of schools also have somewhat different factors associated with variations in their tension levels. For this reason, the remainder of this analysis will use the three categories of school racial composition shown in Table 4.

TABLE 4

Level of Racial Tension
by Racial Composition of School

	School Racial Composition		
	"Black" 5-45% W	"Mixed" 46-75% W	"White" 76-95% W
Tension Mean	26.6	32.3	26.6
Standard Deviation	11.1	15.5	12.8
(n)	32	78	60

For each class of schools, we constructed a long series of regression equations each combining a single predictor variable and one or two control variables. In all classes of schools, school size was used as a control variable. Large schools have more racial incidents, although the actual number of incidents per student may not be greater. In black schools, we further controlled on region because there was more racial tension in the Deep South than in the border states. In the mostly white schools, school racial composition was used as a second control variable.

In Table 5, each line results from a different regression equation. For example, the first predictor is the racial composition prior to desegregation. Among predominantly black schools, there is a coefficient of -.24, indicating that schools which were black before desegregation have less racial tension. We see a smaller coefficient among schools of mixed racial composition, -.08. There was in the sample only one predominantly white school which was black before desegregation. The other lines of the table show the relationship when previous school racial composition is replaced in the equation by other predictor variables. With the small sample sizes, fairly large regression coefficients are needed for significance. In Table 5 we have reported not only the significant relationships but some of the nonsignificant factors which are consistent with the significant ones and lend further support to various hypotheses.

Tension and Alienation

It seems reasonable to begin theorizing about racial tension by asking how black students respond to the school racial climate. Empirically, we know that they are reported as initiating more violence and protesting more about discriminatory treatment in the school. Theoretically, we know that the school system is an institution run by whites, where blacks are sometimes made to feel as if they were somehow intruders. Bear in mind that the predominantly black schools have a 60 percent white teaching staff and that 70 percent have white principals. This leads us to expect a two-step process—social structure leads to black reaction which leads to tension.

Perhaps the simplest explanation, and one that sems to work fairly well, is that racial tension stems from a sense of alienation in blacks. The most common example is the experience of feeling like an unwelcome intruder in a white environment. The absence of this form of alienation may explain why tension is low in predominantly black schools in the sample. Table 5 presents additional data which tend to support this hypothesis. We have already noted less tension in schools which were black before desegregation. We also see less tension when blacks have been assigned to neighborhood schools (line 2), when the school has a black principal (line 4), and when there is a larger black population in the community (line 5). And oddly enough, among predominantly black schools there is less tension when whites are "bused" (line 3)—defined here as attending a school farther from home than necessary. By that definition, 16 percent of the white and 30 percent of the black students in these schools are bused. The opposite of alienation is sense of community or identification. Table 5 seems to indicate two ways to reduce tension by increasing student identification with the school. One is athletics; predominantly black and predominantly white schools with winning football and basketball teams have considerably less tension (line 7). When we began this study, a superintendent in Alabama told me that football was the key element in the desegregation plan. The data seem to bear him out. Another way to develop a sense of community in the school is to take advantage of an attractive physical plant. The data indicate that at least in mixed and predominantly black schools an attractive building is associated with less tension (line 6).

Tension and Reduction of Restraint

The second hypothesis supported by these data is that racial tension results from rising expectations or, more simply, a lack of fear among black students. Blacks in the South have traditionally had few civil rights and been at the mercy of autocratic white police and white adults. Presumably, this has left a residue of hostility, but blacks are unlikely to express their anger unless they feel they can do so without great danger. There are numerous examples, such as the fact that the wave of civil disorders in the 1960s began in the North and West and never penetrated very far into the South. This theory leads to predictions counter both to intuition and to those generated by other theories. For example, black identification with school explains why black schools have low tension; but the "lack of restraint" theory predicts low tension in white schools, which also fits the data.

Table 5 presents other evidence for the "lack of restraint" theory. There is more racial tension where black students are middle class (line 9) and well informed about black history (line 10). There is more tension in communities which did *not* resist desegregation and where the superintendent and school board *supported* peaceful desegregation (lines 12, 11). The more recently desegregation occurred, the less tension in predominantly white and predominantly black schools (lines 13, 14). If blacks have recent memories of fighting for the right to attend desegregated schools, it is likely that they will be more willing to tolerate real and imagined white racism. Finally, there is more tension where the community has a relatively high educational level (line 8). All of this suggests that the more progressive the community and the more self-confident the black students, the more likely that tension will occur.

TABLE 5

Factors Related to Racial Tension

		School Racial Composition		
Source	**Variables**	**Predominantly Black (5-45% W)**	**Mixed (46-75% W)**	**Predominantly White (76-95% W)**
Control Variables				
Principal:	School Size	.50*	.49*	.45*
	School in Deep South	.34		.11
Principal:	% Black of School			
Predictor Variables				
1. Principal:	"Before desegration, was this a white or black school?" (HIGH = BLACK)	–.24	–.08	**
2. Blacks:	"Is there a public high school closer to your house than this one?" (% NO)	–.10	–.23*	–.27*
3. Whites:	"Is there a public high school closer to your house than this one?" (% NO)	.33	–.07	.03
4. Principal:	Principal's Race (HIGH = BLACK)	–.20	–.09	**
5. Census:	COUNTY % NON-WHITE	–.18	–.18	–.17
6. Observer:	Scale–Landscaping; classroom appearance; broken lockers, windows, water fountains; graffiti (HIGH = GOOD CONDITION)	–.18	–.28*	–.07
7. Principal:	Scale–"How did your football team do this school year–was the team undefeated or lost only one game, did they win more than half their games, or less than half? (Repeat for basketball team) (HIGH = BOTH TEAMS HAD WINNING YEARS)	–.32*	.02	–.21*
8. Census:	County Education Level	.32	.34*	.09
9. Blacks:	Scale–Mother's education; family size, homeowner; receive newspaper, own air conditioner; live with both parents (HIGH = HIGH BLACK SES)	.45*	.14	.07
10. Blacks:	Scale–Knowledge of black history figures (HIGH = GREATER KNOWLEDGE)	.15	.17	–.04
11. Leader:	Scale–Superintendent and school board support of desegretation (HIGH = STRONG SUPPORT)	–.07	.06	.21*
12. Leader:	Scale–District, political and business resistance to desegregation (HIGH = LITTLE RESISTANCE)	.16	.11	.13
13. Director:	"In what year did this district desegregate all of its previously white schools, or are some still all white?" (HIGH = EARLIER DESEGREGATION)	.08	–.14	.05
14. Principal:	Year desegregation caused greatest change in racial composition of student body (HIGH = EARLIER)	.06	–.10	.07
15. Principal:	"Has the racial (or ethnic) composition of your student population changed since the 1970-71 school year?" (HIGH = NO)	.01	–.15	.09
16. Whites:	Scale–Mother's education, family size; homeowner; receive newspaper; have air conditioner; live with both parents (HIGH = HIGH WHITE SES)	.01	.14	–.13
17. Whites:	Scale–"Was the elementary school you went to for the longest time–all white, mostly white, mostly black, all black, other?" (Repeat for junior high) (% ALWAYS IN INTEGRATED SCHOOLS)	.05	.00	–.23*
18. Blacks:	Scale–"Was the elementary school you went to for the longest time–all white, mostly white, mostly black, all black, other?" (Repeat for junior high) (% NOT ALWAYS SEGREGATED)	–.02	.07	–.03
19. Principal:	Scale–present students tracked in junior high; 10th grade academic/non-academic tracking (HIGH = MORE TRACKING)	–.23	.17	–.09
20. Principal:	Scale–"Are the student government officers in your school all of the same racial (ethnic) group, or are they from different groups?" (Repeat for cheerleaders) (HIGH = RACIAL MIX)	–.26	.12	–.17
21. Teachers:	"If you have a student biracial committee. . how effective [has it been]" (% SAYING COMMITTEE IS EFFECTIVE)	.29*	–.38*	–.36*
22. Whites:	"How about most of your teachers–how do you think they feel about blacks and whites going to the same school together?" (% WHO SAY THEY DON'T LIKE IT)	.26	.17	.19
23. Blacks:	"How about most of your teachers–how do you think they feel about blacks and whites going to the same school together?" (% WHO SAY THEY DON'T LIKE IT)	.29	–.03	.01
24. Teachers:	% who say most white teachers dislike desegregation minus % who say they like it	.33*	.23*	.04

NOTE: Data were divided by school racial composition and each predictor variable was entered with the control variables in a separate equation, producing a total of 24 equations containing two or three independent variables. Standardized regression coefficients: positive numbers represent more tension; negative numbers represent less tension.

- * $p < .10$
- ** Coefficient not computed; too few black principals or previously black schools.

Tension, Interracial Contact, Racism

The hypothesis that increased racial contacts will tend to eliminate racial problems has often been advanced, but it receives little support from these data. We have already noted that tension does not decrease the longer the school system has been desegregated. We also see in Table 5 little evidence that tension is reduced if black and white students have had experience with integration prior to high school. (There is only one exception—in predominantly white schools tension is considerably less if white students come from integrated elementary and junior high schools (line 17).

The data lend a certain amount of support to the idea that in predominantly white schools the prejudices of white students are an important factor. We notice slightly more tension where white students are of lower SES (line 16). It is in predominantly white schools that violence against blacks is more likely to initiate, so in these schools lower SES whites and whites less experienced with integration may be more likely to make trouble. The "increased contact" argument would lead us to assume the lowest tension where there are approximately equal numbers of whites and blacks since this provides the greatest opportunity for contact. Unfortunately, these are the schools with the highest level of racial tension.

Line 19 shows less racial tension in schools with tracking. Liberals have long complained about the use of achievement grouping to segregate student within integrated schools. We found that achievement grouping by classroom in elementary schools was associated with less racial contact and worse racial attitudes. But in high schools we found the opposite—tracked schools had less racial tension and more positive racial contact. But Alport's (1954) contact hypothesis is concerned only with equal status contact, not with all sorts of contact. There is considerable difference in the average academic performance of black and white students in these schools. It seems to us that heterogeneous grouping in high school subjects blacks to the frustration of being unable to make good grades in competition with whites and helps to convince white students that blacks are stupid. This is another case where the survey data argue against preconceived notions about race relations.

Only in the mixed category of schools do the data seem to support the hypothesis that things will settle down as blacks and whites gain experience. Here schools which are stable in racial composition and have a longer history of desegregation have less tension (lines 13-15). But in general, there is little evidence to indicate that time heals wounds.

Table 5 also presents data to test the hypothesis that racial tension arises as an expression of black frustration with white racism. Here the data are very mixed. On the one hand, in schools with more tension, both white students and other teachers report that white teachers are not sympathetic to desegregation (lines 22, 24). But these data should be taken cautiously since they could just as well mean that in tense schools teachers are blamed for troubles or that blacks are more likely to charge the staff with discrimination and teachers and white students are therefore more sensitive to teacher racist behavior. Moreover, the teacher survey included questions about racial attitudes largely unrelated to school—for example, how teachers felt about living in integrated neighborhoods and how they viewed laws prohibiting racial intermarriage. What we found does not support the idea that tension is a result of higher levels of staff prejudice. For example, in predominantly white schools tension was correlated *positively* with the percentage of teachers opposed to miscegenation laws ($\beta = +.30$). As a final bit of evidence, there is more tension in white and mixed schools with a *larger* percentage of black teachers ($\beta = +.11$ in both cases).

The problem is that the racist staff theory and the "lack of restraint" theory are contradictory. It seems plausible that in predominantly white schools, where blacks are a small minority and likely to be bused and where the staff is relatively unsympathetic, blacks are simply afraid to demand their rights. As more black teachers are added or as white staffs become more accepting, the lid is loosened and tension is likely to increase. Black students should be least fearful in predominantly black schools. They have strength in numbers and are likely to be attending their neighborhood, traditionally black school; they are in a position to rebel against white racism when it is present, especially if they are middle class (line 9). This is supported by the fact that the predominantly black schools are the one place in the data where there is consistent support for the racist staff theory. Here we find a sizeable positive correlation between self-reported negative teacher racial attitudes and tension, exactly contrary to the white schools. The correlation between tension and the percentage of teachers who approve of miscegenation laws is .26 in predominantly black schools.

Notes on a Theory of Tension

Taking the racism and alienation theories on the one hand and the freedom of restraint theory on the other, we generate an hypothesis about unhappy mediums, or falling between stools. On the one stool, we have the infamous tranquility of Southern slavery where there was no racial tension, except what the Yankees stirred up. On the other stool, we have a vision of a future of racial equality where racial tension will be as rare as is tension between Protestants and Catholics now. The problem is that you cannot get from one stool to the other. This helps to explain why racial tension is such a frustrating experience for school administrators, who too often find liberal reforms making things worse instead of better.

Where do these data leave us in a search for a general theory of racial tension in secondary schools? Perhaps the most important thing they do is dissuade us from any search for single factor theory. For example, one is tempted to use a short-term rational model of tension: tension is the response of minority students to direct indication of school racism. While the theory does not fit these data very well, it is also not wrong: the allocation of positions in the student elite to both races tends to reduce tension (line 20). One might also expect a theory based upon racist behavior on the part of white students to be effective in predicting racial tension. In general, this seems not to be the case—not because white students in these schools are unprejudiced, but because white expression of hostility for blacks tends not to take the overt forms reflected in this particular tension scale. It is our view that a school where blacks are the victims of severe prejudice on the part of white students would not appear to anyone, including the black students themselves, as having a high level of racial tension. It would be an unpleasant school, but unpleasant in other ways.

Our best hope for understanding racial tension may be through a general frustration-aggression model, keeping in mind that (1) immediate examples of racial inequality are but one source of frustration which might lead to an aggressive response, and (2) that aggression is not the only response possible to frustration. Frustration in black students may arise from past incidents of discrimination or from non-racial sources entirely. If frustration is present, it may be expressed directly, in violent aggressive behavior; it may be channeled into nonviolent racial protest; it may be inhibited or internalized; or it may express itself in cathartic behavior such as athletics. When two groups have been isolated socially through history, their initial contacts may involve a certain amount of testing behavior to convince each group that a relationship of equality does exist.

This boundary testing behavior may be the most important aspect of racial tension in the racially mixed schools. Recall that in Table 5 schools often failed to fit the general model for white and black schools. In many cases, tension seems to be higher in schools which are successful in handling other aspects of racial relations; we suspect that in mixed schools an increase in racial tension should not be read as indicating a generally unsatisfactory situation. For example, unlike the others, racially tense mixed schools do not have low levels of friendly interracial contact; the correlation between tension and degree of racial contact is very close to zero. Similarly, a successful athletic program or integration of the student elite does little or nothing to lower racial tension.

Favorable factors in the racial climate in mixed schools may simply encourage a higher level of boundary testing behavior which presumably will run its course after a few years of desegregation. Indeed, there is some slight evidence that for mixed, unlike white and black, schools, the longer the school has been desegregated the lower the tension. An alternative explanation, which cannot be tested with this form of analysis, is that tension in mixed schools results from the frustration of lower class white and black students where middle class and other students who can cope with school social relations are highly rewarded. In a situation with good racial contact and a good educational experience for successful students, the unsuccessful ones may become all the more rebellious.

This suggests that those schools which are effective in reducing racial tension have managed to promote symbols of racial equality while exercising firm control on aggression stemming from other sources, have worked to minimize the frustration of adolescence for all students, have provided a variety of outlets for expression of emotion (athletics, music, extracurricular activities), and have worked to develop a sense of community and loyalty toward the school.

Implications

The implications of this model are complex. Just as there is no single theoretical explanation for tension, there is no single cure. Indeed, perhaps the most important conclusion is that a case can be made that tension at a low level, which represents minimal physical danger, is unavoidable and that a policy of reducing racial tension to the exclusion of all else might be a mistake. If a certain amount of racial tension is necessary as a consequence of racial equality, school officials should accept this burden cheerfully. Perhaps the best a school can do is look for constructive, or at least harmless, outlets for the natural anxieties of students thrown into a desegregated situation. The total elimination of tension may have to wait for the next generation or the one after that. This is not to say that the school can do nothing about racial tension. Two entries in Table 5 point to possible aids—working toward elected student leaders from both races and an effective biracial student committee.

The data seem to leave us with five policy relevant recommendations: integration of the student elite, working for an effective biracial student committee, using achievement grouping, strengthening school interest in athletics, and keeping the school plant attractive. These findings are consistent with a general theoretical argument that school officials should work to minimize tension by reducing status inequalities between blacks and whites, by providing constructive channels for the outlet of racial disagreements, and by providing symbols which permit the loyalty of both white and black students to the school community.

This analysis also can be read as suggesting two more radical recommendations. A school administration could disperse blacks widely into predominantly white schools and institute authoritarian, discriminatory policies to "keep them in their place." The data indicate this would work, although few readers would be willing to send their children to such schools. An alternative proposal, which will not appeal to most whites, would be to design desegregation plans in which most of the busing sent white students to predominantly black schools with black principals, leaving the white students in a minority.

This paper can be taken as an argument for both quantitative and qualitative methodologies. At innumerable points, we wished for the field notes of a dozen anthropologists and enthnographers who had observed firsthand the kind of racial tension we were analyzing. Without those notes, the analyst must construct a theory based on some hunches and not very well-grounded hypotheses about schools. Ideally, this survey should have been preceded by fieldwork of an informal nature to gain a better impression of the racial problems in these schools, and it should have been followed by fieldwork in statistically interesting schools, either to gather additional insights to clarify the theoretical argument or to search for hypotheses for those findings which seem inexplicable.

References

Alport, G.W. *The Nature of Prejudice.* Reading (Mass.): Addison-Wesley, 1954.

Becker, H.S., et al. *Boys in White: Student Culture in Medical Schools.* Chicago: University of Chicago Press, 1961.

Coleman, J.S., *The Adolescent Society.* New York: The Free Press, 1961

Crain, R.L., and R.L. York. "Evaluating a Successful Program: Experimental Method and Academic Bias." *School Review* 84:233-54, 1976.

McDill, E.L., et al. *Sources of Educational Climates in High Schools.* Final Report, USOE Contract OE-3-10-080, Department of Health, Education and Welfare, 1966.

National Opinion Research Center. *Southern Schools: An Evaluation of the Emergency School Assistance Program and of Desegregation.* Chicago: The National Opinion Research Center, 1973.

Rosenthal, R., and L. Jacobson. *Pygmalion in the Classroom: Teacher Expectation and Pupil Intellectual Development.* Holt, Rinehart and Winston, 1968.

Stinchcombe, A. "Some Empirical Consequences of the Davis-Moore Theory of Stratification." *American Sociological Review* 28:805-08, 1964.

Suttles, G. *The Social Order of the Slum: Ethnicity and Territory in the Inner City.* Chicago: University of Chicago Press, 1968.

Weber, M. *The Protestant Ethic and The Spirit of Capitalism.* London: Allen & Unwin, 1930.

RACE AND THE "WE-THEY DICHOTOMY" IN CULTURE AND CLASSROOM

Eleanor Leacock
City University of New York

I am going to approach the problem of analyzing race relations in the classroom by exploring certain pervasive habits of thought and action that I call "we-they dichotomizing." By we-they dichotomizing I mean the habit of relating to people in terms of strongly evaluative unidimensional polarities according to which individuals are in essence viewed in terms of higher or lower on a single scale. This style of interaction is grounded in the history of our culture, and it is embedded in the structure of our schooling. It permeates teaching practices and continually defines and reinforces the separation of school children according to indications of status that place them among the accepted "we" in terms of social values or among the rejected "they." To some extent, the pattern is generally Western, but it has taken an exaggerated form in the United States where the institutionalism of racism and the conscious Americanization of immigrants have been central and intertwined historical developments.

The documentation of grossly racist practices in schools and classrooms unfortunately remains important for efforts to democratize schooling, but this paper has another focus. It suggests as important for analysis the covert ways in which well- intentioned teachers defeat their own attempts to succeed with nonwhite and low income classrooms by persistently, albeit indirectly, defining and dividing children along the lines of racial and social status. The paper further points out that the ethic of cultural pluralism, now increasingly accepted as the only viable goal both for our national life and for the world, is conducive to alternative styles of teacher behavior. Description and analysis of we-they dichotomizing versus cultural pluralistic modes of interaction in the classroom can be useful for developing practical educational and curricular materials for teachers who need and want them.

The "We-They Dichotomy" in History and Culture

In a most succinct statement on American attitudes, Conrad Arensberg points out that "twofold judgments are the rule in American and Western life: moral-immoral, legal- illegal, right-wrong, success-failure, clean- dirty, modern-outmoded, civilized-primitive, developed-underdeveloped, practical- impractical, introvert-extrovert, secular- religious, Christian-pagan" (Arensberg and Niehoff, 1968:160). Other cultures have elaborated dual ways of thinking, such as the Chinese Yin-and-Yang and the Zoroastrian dualism of generative and destructive forces. However, Arensberg points out,

> Other peoples do not usually rank one as superior and thus to be embraced on principle (the Christian God), while ranking the other as inferior and thus to be rejected on principle (the Christian Satan). Instead they will tend to rank the two categories as equal and say that each must have its due; or they may not connect them at all with principles guiding conduct.

Arensberg sees the linked attitudes of effort and optimism as important bases for interpersonal evaluation in American culture. "This national liking for effort and activity, and the optimism which holds that trying to do something about a condition or problem will almost invariably bring success in solving it seems to be specifically American," he writes.

> Effort is good in itself and with effort one can be optimistic about success. The high values connected with effort and activity pass quickly to the principle that, "it is better to do something than to sit back and do nothing." When there is an obstacle one should do something about it. Effort pays off with success. This thinking is based on the theory that the universe is mechanistic and man is its master and man is perfectible....
>
> Activist, pragmatist, and moralizing values rather than contemplative, theoretical, sensual, or mystical ones are integrated into the American character (pp. 165-66).

It is noteworthy that one response to the growing national awareness that we must work with what we have by conserving and enriching it, rather than destroying it and expanding into the domains of others, is the fact that some five million Americans apparently participate in groups that emphasize meditation and the search for bodily harmony, themes that are Eastern in inception. A competitive evaluativeness permeates these movements, too, however, though less openly than the intense competitiveness that characterizes our occupational structure and our schools. To return to Arensberg:

> Serious effort to achieve success is both a personal goal and an ethical imperative. The worthwhile man is the one who "gets results" and "gets ahead." A failure "gets nowhere," or "no results" for success is measured by results (though there is some "credit for trying"). The successful man "tackles a problem," "does something about it," and in the process "gets ahead." His success is measured in terms of his positive solution of the problem. A failure is unsuccessful through his own fault. Even if he had "bad breaks," he should have "tried again." A failure in life "didn't have the guts" to "make a go of it" and "put himself ahead."
>
> This is a very severe moral code.... It calls all those in high positions successes and all those in low ones failures,

> even though we know that there is "a need for Indians as well as chiefs" (p. 166).

And, I would add, even though we cannot help but know how unequally the possibilities for success are distributed.

My point is that the emphasis on identification as "we," the successful, as opposed to "they," the failures, is bound up in our history and ideology with the definition of: first, "we," the whites, and "they," the non-whites, Indians and Africans, then Asians, and more recently peoples of Hispanic cultures; second, "we," the Americans, and "they," the foreigners and immigrants, the "un-Americans"; and third, "we," the "middle-class," the "solid citizens," and "they," the poor, the lower classes, the marginal workers, the unemployed. These definitions are part of what are loosely referred to as WASP or middle class values, and their historical development can be traced. For instance, the historian Edmund Morgan (1975) has documented in fine detail the process whereby traditional European attitudes of class snobbery were redefined in terms of color, when color was made the mark of actual or potential slavery and color caste was formalized as a central feature of American social-economic structure.

It was taken for granted that the great wave of European immigrants in the nineteenth and twentieth centuries were to "become" Americans–that is, that they were not just to live and get along in the United States, but virtually change their cultural identity in a very short time. Yet it is a rather extraordinary idea to "become" another nationality. One would not, for example, consider it possible to become French, no matter how long one might live in France or how well one might learn to speak French. However, immigrants were expected to become Americans, or at least to raise their children to speak only English, and to adopt American food and living styles.

Schooling in urban centers was geared to the making of Americans, and out of this arose a major anomaly of American national life–the stress on cultural conformity in the face of, and because of, great cultural heterogeneity. The value placed on monolingualism epitomizes this paradox. In this most diversified of nations, bilingualism is considered a handicap. Everywhere else in the world (with the exception of England), bilingualism is an advantage and often a necessity. I remember hearing a Chinese boy recite a poem in a fourth grade classroom. While fluent in English, the boy had a slight accent–in fact, a rather charming lilt, for in Chinese tonality has phonetic value. When he finished, the teacher said to me in a whisper audible throughout the classroom: "They speak Chinese to him at home. Isn't it terrible?" After thus derogating a perfectly bilingual child in front of the class, she told him in a supposedly supportive tone, "All right, and next time you'll do it better." Yet this was on the whole a good teacher, known as one of the best in her school.

The process of becoming American, then, meant becoming sensitized to those attributes of personal and linguistic style by which people were assessed as they competed for economic well-being and occupational security. The so-called open class system of a frontier country challenged personal abilities and initiatives, but it also robbed people of a certain security derived from the certainty of traditional occupational status. Unremitting competition was the order of the day. And, of course, class mobility was in fact restricted. A new aristocracy formed at the top, and most of the mobility that took place was cyclical. In the course of a generation or two, European craftsmen regained the relative status they had given up in leaving their original countries; and children who experienced as upward mobility their parents' rise in job seniority and economic well-being did not interpret as downward mobility their own starting over when they became young parents. Furthermore, the increasing availability of consumer goods at reasonable prices that accompanied industrial development led to an escalating game of "keeping up with the Joneses" in which one validated individual effort with "success" without changing one's relative status.

Thus the pattern became set, as it still remains despite emergent themes in American culture more in tune with contemporary realities. Constant attention is paid to culturally prescribed attributes of status as part of a continual effort to achieve a modicum of upward mobility and economic security. Always at the bottom, however, the nonwhite population serves to assure whites they are at least better off than someone else. Color is critical in defining "they," although it is thoroughly intermixed with concepts of class and can be compensated for by attributes of class status if these are unequivocal enough.

"We-they" dichotomizing as a way of life involves downgrading the "they" as much as upgrading the "we." In the definition of class attributes in speech styles and manners that is central in the socialization of children, the things one should *not* do if one is to become (or remain) part of the successful and worthy "we" are often more strongly defined than those one should do. Children whose parents are "in" are called on to validate their status by their behavior and performance; children whose parents are culturally defined as "they" are faced with a bitter predicament. In either case, the continual drawing of models that pervades curriculum content and teaching styles, both explicitly and implicitly, is anything but conducive to intellectual development. Jules Henry (1960:274), who so brilliantly described for middle class schools socialization practices relevant to this discussion, wrote caustically:

> Nowadays, in America, there is much talk about teaching children to think. In five years of observation in American schools, however, we have found very little behavior that tends in this direction... Thinking would seem to involve an analytic process of some kind and also a process of synthesis. Almost none of this takes place in elementary school (although we have found it occasionally) and little more even in high school science courses....

In *The Lonely Crowd*, Riesman, Glazer, and Denny (1953:82-84) were also referring to middle class schools when they wrote of the teacher's role as "that of opinion leader," in the "socialization of taste and interest" that underplays "the skills of intellect" and overplays "the skills of gregariousness and amiability." They described the teacher as conveying "to the children that what matters is not their industry or learning as such but their adjustment in the group, their cooperation, their (carefully stylized and limited) initiative and leadership." John Holt, who has written insightfully about the fearfulness and intellectual constriction of children in middle class schools, has become thoroughly discouraged about possibilities for reform and says so bitterly in his latest book (1976). Meanwhile the literature on ghetto schools describes in angry detail the process whereby black children are taught not to learn or at least not to learn much of what schools are supposed to teach.

Teachers devote themselves with the best of intentions to their part in the socialization process. They learn that their task of making good citizens of children calls for recognizing those who will succeed and those slated for failure, and they simply are going along with the culture of school and society when they do this on the basis of accepted indices of social status. For example, Rist (1970) documented the commitment of the teachers he studied, black women in an all-black school, who concentrated their efforts on the higher status children and tried to insulate them from the bad influence of those already designated for failure at the kindergarten to second grade level.

Cultural Pluralism as a Goal

Where, then, are the sources for change? I think the most positive development in relation to short-range and at least partial reforms in schooling is the assertion of cultural pluralism as a desirable goal. Cultural pluralism challenges a single standard for evaluating children. True valuing of cultural differences is inseparable from true valuing of individual differences, and appreciation of diverse individual potentials is necessary if the educational principles advocated in teacher training are ever to be applied.

Admittedly, a formal commitment to the value of differences is commonly made in school, especially during "brotherhood month." However, this commitment is typically phrased as "they" are really same as "we" are, though they may not seem to be, or as "they" are just as good as "we" are. There is no challenge to a unilineal scale according to which people are evaluated and according to which the teacher and those included as "we" are eligible for the higher ranges. The interest and excitement that could accompany learning about differences is submerged by the concern with relative merit. Differences loom as problems. They are sensed as threats, for people cannot just be different—someone has to be "right" and someone "wrong." Differences are to be "tolerated," not enjoyed, except in superficial compartmentalized ways, such as when viewing national dances or envisaging travel to colorful places.

In recent years, the goal of genuine cultural pluralism has become something more than a humanistic and aesthetic statement or a theme in introductory anthropology. It has taken on real embodiment in the contemporary world, both nationally and internationally, as Third World nations abroad and minority groups at home attempt to achieve economic and political equity. The goal of cultural pluralism as expressed today stresses ideological autonomy and the full valuing of one's own history and traditions. It also questions Western patterns of urbanization and industrialization as the model all others should follow in the process of economic development. Cultural pluralism as an ideal is certainly not without great contradictions and confusions—but that is true of any broad historical process.

It is important to recognize that cultural pluralism is not contrary to integration but essential for its realization. Without self- respect and mutual respect, integration means no more than the assimilation of the socially discriminated against group into the dominant one or, in effect, the acceptance by the former of a subordinate status. The "we-they dichotomy" is born of assimilation. Its patronizing addendum, "they are *really* as good as we are," means that "they" can and should become like "we." While never very salutory, such an orientation is thoroughly anachronistic today when the West, a cultural innovator since the Industrial Revolution, needs new models for living.

The significance for schooling lies in the challenge that cultural pluralist goals make to narrowly ethnocentric and status linked criteria for evaluating children and their performance. Thus, black parents have sought to influence their children's education, and monolingualism has been challenged by Puerto Ricans who have asserted their intention to be bilingual. Cultural uniformity as a national aim of education also has been challenged by native American groups who insist upon the right to oversee the education of their children. Indeed, many American Indian peoples have demonstrated the reality of cultural pluralism by maintaining their identity as culturally distinct enclaves for centuries while at the same time endeavoring to participate without discrimination in the larger society, according to individual abilities and interests.

As an example of cultural pluralism in education, Vera John compares a school for Indian children with progressive schools for affluent children. John (n.d.) writes:

> Schools which support active and functional learning in children in a setting which is rooted in their community do not produce failure. Two very different kinds of schools, both exceptional, come to mind. One is the experimental, comfortable, friendly, non-competitive school—usually private—which services upper-middle class children. The teachers are called by their first names; play and learning are woven together; and the children are looked upon as capable and exciting. Those who can read get new books, and those who cannot are not pressured. The goal is universal literacy, but the time-table is determined by the children.
>
> In a very different setting, a Pueblo kindergarten class along the Rio Grande, I saw a group of children as secure, active and comfortable with themselves and adults, both teachers and visitors, in their classroom as the children above. Their home-made books depicted a story of an abandoned Pueblo house; their teacher appeared in their books as a ghost for Halloween; the room was full of their paintings; and they learned their numbers by charting their weight gain. The school is in the middle of the village; parents come in; workmen (Pueblo, Navajo and Anglo), who are building an additional classroom, drink their coffee in the classroom and the children imitate their digging and building during their out-door play.
>
> This community is rooted in the long and continuous history of the Pueblos; they treasure their culture. At the same time, they have effectively developed new economic programs which have resulted in a higher standard of living for the entire Pueblo. The children are well-fed, comfortably dressed, but although many of them have been to other towns and cities, none of them know luxury.

Although widely divergent, the two types of schools share crucial features. In both the children are respected, and in both they are expected to learn. In both, in John's words, "the children are able to learn in ways which do not conflict with their previous experiences."

Qualitative Research and Theoretical Formulation

As I shall detail shortly, there are a number of areas in which research could help to identify, document, and analyze effective teaching in inter-group settings. Qualitative research is clearly appropriate for such purposes because it permits the development of models based on concrete examples from realistic classroom settings.

I share the opinions expressed by others at this symposium that there are different ways to define qualitative and quantitative and that they should supplement each other or be combined according to research purposes. In a comparative study of four schools in neighborhoods that differed by income level and race, the orientation of my co-workers and myself was in general qualitative, but of course we used quantitative data and techniques for selecting sample schools, as well as simple scaling and coding techniques for counting teacher-child interactions of different types and for rating teacher attitudes in various ways (Leacock, 1969). But because we always stayed close to our original observational and interview materials, we were able to make maximum use of an extremely important type of datum—the key incident.

Key incidents epitomize underlying relationships that quantifying methods may suggest but seldom directly reveal. They help explain statistical correlations. We found that in the middle income white fifth grade in our sample, the children toward whom the teacher felt positive had an average IQ score some eleven points higher than those toward whom she felt negative. In the low income black school, the reverse was true; the teacher felt positive or neutral towards children whose average IQ score was almost ten points lower than those about whom she felt negative.

This second teacher, asked in an interview about the kinds of things she felt her pupils should be getting out of school, stated: "First of all discipline. They should know that when an older person talks to them or gives a command that they should respond, that they should listen...." The teacher, a black woman, was by no means a stern disciplinarian; her goals for black children from low income homes reflected cultural prescriptions. Her key statement pointed up the critical role of schooling to black low income students in the context of the total social-economic structure—i.e., to train them to take orders in low status jobs or cause them to drop out and become future "unemployables."

Thus when it came to characterizing what we felt to be the most cogent differences among classrooms, it was often key incidents and teacher statements that best summarized central messages being conveyed to the children. A teacher in the middle income white school made her behavioral demands by saying: "I will choose two lovely children to show their book reports to our visitors. I will only choose two of the nicest people, the two with the best self-control." By contrast, a teacher in the middle income black school who was lecturing her class about self-control said, "Now you've had many compliments, but I think we need to stop once more and ask, is this the *best* we can do?" In the first classroom, the "nicest" children were to be rewarded. The statement in the second classroom seemed a parody on the demands and restrictions placed on black people if they are to compete successfully in what is to them a highly restricted middle class arena. To be recognized as good is not dependable; one always has to be *better*.

Such incidents ring true in relation to what we know about our society in general as well as in relation to the classroom data we were analyzing. However, their selection was also guided by a theory of schooling as involving a fundamental set of social, economic, and political relationships. As I see it, the basic methodological problem, whether the orientation is qualitative or quantitative, is always how to clarify social phenomena in terms of relationships and relationships among relationships, rather than dealing with them as essentially static characteristics that intercorrelate. The latter leads to biological reductionist distortions. These constantly creep into research designs and methods, especially in view of our strong cultural myth that inborn psycho-biological characteristics determine social patterns of behavior and our metaphysical habit of viewing reality in terms of separable qualities or Platonic essences rather than interactive processes.

Levels of integration theory affords an important corrective to biological reductionist formulations. The levels of integration concept has been elaborated by a number of biologists (Szent-Gybörgyi, 1966; Redfield, 1942; Novikoff, et al., 1945; Tobach, 1976). They point out that matter is organized or integrated in progressively more complicated levels as one moves from the physical (atomic and molecular) levels, through the biological or physiological levels, to the social level and that new properties kept emerging as "higher" forms of matter evolved from "lower" ones. Each successive level is based on properties of "lower" systems but functions according to properties specific to its own level. In one sense, the functioning of the digestive system, for example, is no more than the sum total of the molecular movements that make it up. However, its origin and functions can no more be explained or understood in terms of its constituent molecules than the movements of the planets in our solar system can be explained in terms of the molecular movements of which they consist.

Similarly, social life has laws of its own that define patterns of social behavior. In a superficial sense, a society is no more than the sum total of movements (behaviors) of its constituent individuals—hence the appeal of biologically reductionist theories that interpret social phenomena in terms of individual psychological characteristics. However, only behavioral universals can be explained in such terms: eating, sleeping, laughing, or feeling sorrow or anger. Socially differentiated behavior, which is virtually all actual behavior—that is, when, how, why, and how much people eat or sleep or feel anger—can only be explained in terms of *social* processes or patterns of human *interaction*.

The concept of "territoriality" from the field of ethnology is an example of metaphysical or "typological" (Mayr, 1959) and reductionist formulation. All animals must distribute themselves in space and have evolved patterned ways of doing so. Some animals mark areas and keep others of their species away. Most animals do not, but simply spread out in one or another kind of grouping in the search for food. Inquiry into the bases on which different animals do this is hindered by the tag "territoriality," conceived as a given, an "essence," that animals have in specifiable amounts. The tag lumps adaptive behaviors that have evolved in many different ways and obscures relationships of animal species with each other and with their environments.

In the study of race relations, the simplest form of biological reductionism is direct racist allegation of deficiency in some socially valued trait. Studies have continually attempted to demonstrate racial inferiorities and have been rebutted, only to emerge again. Otto Klineberg's (1935) classic study of rising IQ scores with improved schooling would probably have ended the matter once and for all if the issue were purely scientific and not economic and political. The virtual consensus among anthropologists that there is no basis for assuming group differences relevant to effective social functioning derives, I think, not only from cross-cultural knowledge and a culturally relativist theoretical perspective, but also from the fact that anthropologists in field research put themselves in the position of learners. Anthropologists learn from rather than evaluate, boss, service, or otherwise manipulate people who fall into the category of "theys" in terms of social status. They thereby learn to recognize and respect intellectuality among low status people, a quality that typically goes unnoted by those in socially superordinate positions.

However, the "culture of poverty" concept, derived from anthropology, exemplifies the way in which a set of social relationships can become reduced to an attribute, a quality of deficiency some children possess. The interlocking structures of urban institutions, including occupational opportunities and real estate interests as they mesh with the structures of schools and neighborhoods, confront poor and especially black poor children with a repeated series of problems. It is in the nature of the institutional structure that a limited number of these problems can be overcome by a limited number of these children; everybody, or even very many, cannot become "middle class." However, the complex set of social relationships involved has been translated through the "culture of poverty" tag into an entity characterizing children, not the society.

By the same token, to seek measures of teaching effectiveness in terms that imply some specific quality a teacher possesses in greater or lesser amount denies the complexity of the teaching function in our society and the fact that teachers represent a set of relationships that are in considerable measure beyond their control. This is not to say that in a practical sense some teachers are not better than others, nor to suggest that they should not as individuals be held responsible for doing their best in whatever situation they teach. In fact, when poor parents call teachers to account it helps change the set of relationships inimical to successful teaching. However, "teaching ability" is not an inherent quality of a teacher, but a certain point in the accumulated set of relations in which the teacher has been and is involved. Hence the elusiveness of teacher assessment. While a few extraordinay people stand out, most teachers are reasonably successful with some subjects and not others, with some children and not others, with some grade levels and not others, and so on.

Some two decades ago, it was thin going when one looked for systematic documentation of differences in schooling according to the racial and class status of pupils or for research that showed teaching styles to be as strongly patterned by differential expectations of and attitudes towards students as by the personality or educational orientation of individual teachers. Subsequently, studies of many kinds—quantitative and qualitative, personal accounts and formal observations, bitter criticism and dispassionate analysis—have made clear how constraining is the network of social-economic and political relations within which teachers and principals must function. Case studies of ghetto schools, statistical studies of school performance and its correlates, structured studies of differential teacher behaviors, institutional and historical analyses of the educational system as a whole, critiques of teaching methods even in "good" schools, and third world critiques of schooling and its social implications—each adds a different dimension to the analysis of schooling as differentially training children for different stations in society.

One might throw up one's hands in hopelessness at the whole picture were there not also studies of successes where commitments to change have been made and were there not parents and educators who keep on trying. After all, attempting to change the schools is part and parcel of continuing attempts, despite setbacks, to democratize and equalize our society generally. Research and documentation not only can clarify where the greatest leverage for change may lie but also can help foster an interest in and commitment to school reform.

Suggestions for Qualitatively Oriented Research

The "we-they dichotomy," embedded in curriculum materials and teaching styles, strikes back on the classroom level at the institutionalization of differential education and socialization for children of different racial and class backgrounds. The Deweyan principles of respecting children's ability to learn and building teaching on children's experiences are familiar to educators, but putting them into practice is another matter. It is particularly frustrating for educators who try to apply these principles to confront the stubborn persistence of antagonistic teacher-pupil relations that follow from the structure of schools and school-community relationships with regard to institutionalized racism. My suggestion is that it can be useful to analyze and document both divisive and undermining techniques unwittingly used by teachers who are successful in heterogeneous and minority classrooms.

What might be some useful areas for analysis? One important area is the ways teachers differentiate children. Generally, teachers differentiate through: (1) selection of reading and other groups, selection of officers, monitors, and the like, and seating arrangements (c.f. Rist, 1970); (2) selection of materials to post on classroom walls (in one low income black classroom I studied, only the names of children eligible for free lunch were posted); and (3) direct references to specific children in overt model-setting statements as well as the myriad of direct and indirect instructions and reactions concerning children's work and behavior. Silberman (1971) found that in middle income classrooms, teachers used especially favored children as role models; in an extremely well-constructed study, Hartley (1972) found that teachers were "inordinately critical" with low income children, "often giving negative feedback to pupils for behaviors ordinarily regarded as appropriate."

A second area for analysis of how "we- they" definitions operate in the classroom concerns the utilization of children's experiences. Textbook denial of the existence and/or worthiness of children who are nonwhite and poor is destructive. This denial is commonly reinforced by teachers' negative responses to the experiences these children proffer.

In the middle income classrooms I observed, teachers often made intellectually superficial responses to children's discussion of personal events, but they were at least supportive and children were rewarded for their contributions. By contrast, during a session on transportation in a lower income black second grade, a boy talked at length and excitedly about the planes he had seen on a visit to the airport. When he finished, the teacher ignored the rich content of his tale. Instead, her curiosity predominated. Since "culturally deprived" children are not supposed to go anywhere, she asked, "Who took you?" The boy, nonplussed, said, "Day care." Her stereotype confirmed, the teacher said, "Oh," and moved to another topic.

Later, when a girl told about seeing her father off on a train trip, the teacher contradicted her, saying it was her uncle, not her father. The girl tried to argue, then gave in and sat down, silent and confused. Then when the class shifted to reading, the teacher asked the children if they wanted to work on a certain story. In union, they responded, "No." In a beguiling tone, the teacher suggested another story; again, the answer was, "No." The contrast to the eagerness with which children in the middle income classrooms complied with and tried to second guess the teacher was striking. It would be easy for a teacher to conclude that black children are, after all, fatherless, unmotivated, culturally deprived, hard to reach. What the record showed, however, was that the children were responding to the teacher with the same denial that she had just extended to them.

Teachers' failure to build on children's experiences often flows from lack of knowledge and from an unwillingness to put themselves in the position of learners from nonwhite or poor children. It would be helpful to document the kinds of knowledge children from different backgrounds have so it could be incorporated into lesson plans that would strengthen a child's sense of competence. When conducting research on the largely new and most impressive Zambian school system, I recorded many out-of- school activities of children that were similar to activities prescribed in teaching manuals for experimental schools. However, British advisors and elite Africans were devising lesson plans as if working class children had no experiences of their own on which to build. For example, although children in Zambia play a simplified version of a ubiquitous checker-like game, learning to assess moves by rapidly adding and subtracting small numbers, I did not find teaching manuals that used examples from the game to illustrate mathematics problems.

Teachers' negative reactions to the experience of black and poor children arises from an additional source, one that is complicated to handle since it concerns the superficiality of the curriculum generally—what Jerome Bruner (1959) has criticized as the "pablum" fed school children. Children of the poor are not sheltered from social realities to the degree that affluent children are. Hence they violate a norm of school culture—that only the "nice" should be brought into the classroom and that anything ugly or controversial must be avoided. Poor children, by their very knowledge of the world, identify themselves as the rejected "theys" and are made to suffer. As an example, Herbert Kohl (1966:27) cites two poems written by eleven-year-old girls:

Shop with Mom

I love to shop with mom
And talk to the friendly grocer
And help her make the list
Seems to make us closer

The Junkies

When they are
in the street
they pass it
along to each
other but when
they see the
police they would
run some would
just stand still
and be beat
so pity ful
that they want
to cry

"Shop with Mom" was highly praised and published in the school paper; the other poem was met with horror and put aside. Though wise about and sensitive to a major social problem, it violated the taboos of the classroom. There has been at least a limited change in some urban areas since Kohl's writing, but the myth of the "nice" world still constricts teachers who would wish to broaden their curriculum and calls for analysis and documentation of its effects on children's learning.

Another area for research is relations between teachers and parents, which educators and researchers alike recognize as mediating the attitudes of teachers towards children. We-they dichotomizing essentially flows from the social distance between teachers and parents since it defines the distance between teachers and children. I was introduced to this phenomenon when as a parent I somewhat unwillingly became active in a PTA squabble. Almost immediately, my eight-year-old daughter's role in her classroom was transformed. Instead of saying sadly that she had no friends at school, she became part of a "social set," a "somebody" to both pupils and teacher. I realized that in school terms, I had been seen as belonging to the category of "working mother of a large family who neglects her children." I had four children and purposely did not check with the six- and eight-year-olds about their homework or do more than they asked of me. I learned that they were thereby being "deprived" in contrast with the other middle class children in a somewhat heterogeneous public school where parents were expected to be teaching aides. When I was shifted to the category of "professional family whose children will succeed," the ways in which my children subsequently were favored were at times embarrassing.

In a report on relations between white teachers and black and Puerto Rican parents, Anne Okongwu (1975:13-14) states the teachers' feelings that "they found it difficult and frustrating to teach" the children in their classes and that "they didn't have very high hopes for these children in the future." Okongwu writes:

> They consistently [stated]...that the student's lack of academic progress was not a result of any failure on their part but rather the result of students' disruptive behavior in the classroom, their unwillingness to learn, the crowded conditions of triple session, lack of parent interest and cooperation, poor home conditions, family structure or innate low intelligence. Some of the teachers emphasized the above by verbalizing negative feelings about the families of some of the children in their classes and stated repeatedly that they got no "cooperation" from the parents...None of the teachers, however, suggested that they were inadequately prepared to teach black and Puerto Rican students from low socioeconomic backgrounds or that they lacked the tools to adequately perform their jobs.

In this instance, a step toward more positive teacher-parent relationships was made by inviting parents to a morning coffee hour. Rather than a mimeographed announcement of a PTA meeting or an implicit command to come to discuss a problem, this was a somewhat social invitation appropriate for equals. The response was most positive, providing a basis for what could be the next step in such a program—developing a fuller dialogue between teachers and parents about educational problems.

These are only a few examples of areas in which "we-they" dichotomizing is manifested in classrooms and which could be documented and analyzed through qualitative research. Studies that yielded concrete classroom examples could help teachers shift towards broader, more realistic, and more positive role definitions than are typical—towards models with less of a moralistic emphasis on behavior as such and more of a supportive emphasis on what children from nonwhite and working class homes have to offer to their own educational process.

References

Arensberg, C.M., and A.H. Niehoff. *Introducing Social Change: A Manual for Americans Overseas.* Chicago: Aldine, 1968.

Bruner, J.S. "Learning and Thinking." *Harvard Educational Review* 29:3, 1959.

Francis-Okongu, A. "What Poor Children Are Learning for the Future." Paper delivered at the 74th Annual Meeting of the American Anthropological Association, November, 1975.

Hartley, M.R. "Differential Treatment of Upper-Income and Lower-Income Children by the Public Schools" Unpublished dissertation, University of Oregon, 1972.

Henry, J. "A Cross-Cultural Outline of Education." *Current Anthropology* 1:269-305, 1960.

Holt, J. *Instead of Education.* New York: E.P. Hutton & Co., 1976.

John, V.P. "Transforming the Structure of Failure." Ms., n.d.

Klineberg, O. *Negro Intelligence and Selective Migration.* New York: Columbia University Press, 1935.

Kohl, H. "Children Writing: The Story of an Experiment." *The New York Review of Books,* Nov. 17, 1966.

Leacock, E.B. *Teaching and Learning in City Schools.* New York: Basic Books, 1969.

Mayr, E. "Darwin and the Evolutionary Theory in Biology." In B.J. Meggers (ed.). *Evolution and Anthropology: A Centennial Appraisal.* Washington, D.C.: The Anthropological Society of Washington, 1959.

Morgan, E.S. *American Slavery—American Freedom: The Ordeal of Colonial Virginia.* New York: Norton Press, 1975.

Novikoff, A., et al. "The Concept of Integrative Levels and Biology." *Science* 101 (No. 2618), 1945. Rebuttals and reply, *Science* 101 (No. 2632), 1945; *Science* 102 (No. 2651), 1945.

Redfield, R. (ed.). *Levels of Integration in Biological and Social Sciences.* Lancaster (Pa.): Jacques Cattell Press, 1942.

Riesman, D., N. Glazer, and R. Denny. *The Lonely Crowd.* Garden City (N.Y.): Doubleday Anchor, 1953.

Rist, R.C. "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education." *Harvard Educational Review* 40:411-51, 1970.

Szent-Gybörgyi, A. "Fifty Years of Poaching in Science." *Graduate Faculties Newsletter,* Columbia University, March, 1966.

Silberman, M. "Teachers' Attitudes and Actions Toward Their Students." In M.L. Silberman, (ed.). *The Experience of Schooling.* New York: Holt, Rinehart & Winston, 1971.

Tobach, E. "Evolution of Behavior and the Comparative Method." *International Journal of Psychology* 1976.

CRITIQUE

Gladys G. Handy
Pennsylvania Department of Education

Increasingly it seems that bureaucrats are complaining that the findings of educational researchers are of very little use while, at the same time, researchers are complaining that policy makers ignore the results of their hard work. Why might this be? As an approach to critiquing the papers by Crain and Leacock, I'd like to share a few reflections on this issue that come out of my experience as a bureaucrat.

From my perspective, one part of the problem is in the selection and definition of research questions. In

Crain's paper, for example, the definition of racial tension relates primarily to tensions perceived by whites—to the problems whites may have with black behavior. The tension instrument does not in my view sufficiently take account of the tensions blacks feel. Yet I know very well that in order to understand racial tension in schools and to make appropriate policy decisions, I need information about the problems as perceived by both blacks and whites. Protests and outbreaks of hostility are usually a last expression of problems that have gone before, and unless research gives me a balanced picture of these problems, I will tend not to use it.

Similarly, in neither paper does the definition and discussion of prejudice satisfy my needs as a policy maker. Crain's definition of prejudice does not reach the subtle dimensions of behavior that Leacock points out. Yet while I am very much taken with the idiosyncratic, anecdotal exposition of prejudice in the Leacock paper, it worries me that I have no information about how widespread the behavior she pictures is. I can't generalize, and I can't make distinctions between what is idiosyncratic and what is amenable to policy decision.

Neither paper quite satisfies my needs for information with which to make policy or implement programs. At what might be called a macro level of educational policy making—that is, the level of a state or federal agency—I can influence budget, regulations, and legislation and I can monitor activities at the micro or school level. Given this role, there is very little I can do with the findings in either paper. On the one hand, the paper by Crain leaves me with the feeling that there are important issues associated with racial tension in schools that were not addressed by the study—issues with which I, as a policy maker, must deal. On the other hand, the Leacock paper points out many of those issues but leaves me with the pronounced feeling that there is not a thing I can do about them.

Part of the problem, perhaps, is that at the policy level, we have integrated the qualitative and quantitative approaches to collecting educational information. We cannot do our work without looking at numbers and statistical data in general. This very fact makes us keenly aware of the problems of obtaining correct data and gives us a healthy cynicism about what those data mean. At the same time, we also are engaged in a kind of mini ethnography when we look in more depth at specific schools and classrooms—a process which teaches us to be cautious in making policy decisions at the macro level based on information from a very limited data base. At least part of the policy maker's failure to use the results of educational research, then, is because so much of the research relies on limited methodology—for example, only statistical or only ethnographic—when what we want and need is information collected through an integrated approach.

Still another problem from the perspective of the bureaucrat is the general failure of educational research to consider the impact of its findings. Crain's paper, for example, presents a "policy relevant finding" that there is less tension in high schools with winning football and basketball teams. What is the impact of the "policy relevant finding" on policy? In the first place, not all high school athletic teams can be winners! In the second place, and less facetiously, there are potential conflicts at the policy level in both objectives and budget. We might well think it important to decrease competitiveness between racially tense schools and our budget priorities might not allow the investment of money to support a highly competitive inter-school athletic program.

My fundamental recommendation, then, is that educational researchers need to be more sensitive to the needs of policy makers. Researchers should ask themselves, "Who is the ultimate user of my findings?" and frame their research questions and plan their research design and methodology with an awareness of that audience. And either through research reports themselves or through some additional developmental process of bridging between research and the operational level, the impact of research findings must be examined. Only in this fashion, I believe, can educational research such as that presented by Crain and Leacock influence the behavior of policy makers in education.

CRITIQUE

Robert E. Wentz
Saint Louis Public Schools

In critiquing the papers by Crain and Leacock, I began from the vantage point of a practicing public school administrator, not a researcher. My concern is with finding in the work of researchers information that will help educational decision makers—primarily teachers and administrators—have a more objective understanding of the "real world." It is not enough simply to compare the advantages or disadvantages of qualitative and quantitative methodologies. As Rist so ably points out in his paper for this conference, such comparative analyses reduce "the complexities and nuances of research approaches...to simple and rigid polarities" and obscure "the dialectic and interaction among all efforts to 'know' or to 'understand.'" With a concern for knowing and understanding in order to effect meaningful change *in the classroom,* I obviously agree with Rist that we "only hinder and cripple ourselves by a continued fixation upon what is 'good' about one approach or 'bad' about another."

It is my basic premise that both quantitative and qualitative methodologies are essential in the exploration, discovery, and refinement of knowledge about such critical concerns as race relations in the classroom. While Crain and Leacock each are writing from the perspective of either a quantitative or qualitative educational researcher, they both acknowledge the role of the other's approach. Thus, Crain stresses: "At innumerable points, we wished for the field notes of a dozen anthropologists and ethnographers who had observed firsthand the kind of racial tension we were analyzing. Without those notes, the analyst must try to construct a theory based on some hunches and not very well-grounded hypotheses about schools." Leacock also supports the use of both methodologies when she says that "they should supplement each other or be combined according to research purposes."

In my view, qualitative and quantitative methodologies should follow each other in educational research and not "be combined" as Leacock seems to suggest. To use both approaches simultaneously may result in mucking around with data (post hoc analysis) and lead to inferring causality instead of "possible relationships." In fact, a most realistic approach would seem to call for qualitative exploration to establish testable hypotheses, application of quantitative methodology to test those hypotheses, and then use of qualitative methodology again to investigate causality.

Among those of us attending this conference who are in one capacity or another practitioners rather than academicians and researchers, the call for the use of both qualitative and quantitative methodology in educational research has been repeated. We hope our words will not go unheeded. Rist summarized it all rather succinctly when he said, "No one methodology can answer all questions and provide insights on all issues."

Turning now from the general subject of the merits of using both approaches in educational research, I would like to make a few specific comments about each of the papers, beginning with Crain's report on his survey of racial tension in southern high schools. I'd like to make five points.

First, I question the validity of the racial tension instrument. Items 1 and 7 on Table 2, for example, both measure perceptions of whites and blacks, respectively, about "problems between blacks and whites in the school." If the instrument were valid as a measure of racial tension, one would expect these items to be highly intercorrelated. The fact that they are not (see Crain's Table 3) says to me that the instrument may not have internal consistency. One would need to see the original data by racial composition of schools to infer further problems.

Second, I think it is important to recognize that responses to such items as "complaints of favoritism toward whites" or "black attacks on whites" take on a different contextual meaning in predominantly white schools than in predominantly black schools. To generalize to all settings from these responses which were made in specific contexts is a bit dangerous.

Third, it appears that Crain gives unwarranted significance to the data in his report about the behavior of school superintendents and communities during the initial stages of school desegregation. *Only four* telephone interviews with community leaders were used to collect these data—which does not speak well for survey methodology!

Fourth, Crain's categorization of high schools by racial composition is of considerable interest. I am not sure why the categories were defined as they were; to me, for example, "mixed" means exactly that and not 46-75% white as defined in this survey. It appears that one would have different findings if the categories had been defined differently—as, for example, 5-35% white, 36-65% white, and 66-95% white. In any event, the data certainly suggest that studies on desegregation should look closely at the racial composition of schools and not attempt to generalize about desegregation without reference to the proportions of black and white students.

Fifth, Crain made a number of statements implying that the data were similar for all three categories of schools. Yet the coefficients reported in Table 4 indicate that there were important differences when racial composition is examined. He comments, for example, that "there is more racial tension where black students are...well informed about black history (line 11)"–a statement that his data indicate is true for predominantly black and for mixed schools but not for predominantly white schools.

The same need to qualify conclusions by the racial composition of schools also applies to the author's five "policy relevant findings" which he bases on five predictor variables associated with lower tension. Examining the data in Table 4, we find that lower tension was not associated with any of these five predictor variables across *all three* categories of school racial composition. Thus, while there was low tension in predominantly white and predominantly black schools with winning athletic teams, tracking, and student leaders from both races, this was not true in mixed schools. Likewise, tension was not low for the other two predictor variables in one of the three categories of racial composition. Tension was not low in predominantly black schools where there was an effective biracial student committee nor in predominantly white schools with a well-maintained and attractive physical plant.

It is my interpretation that the data are neither strong enough nor consistent enough across all school compositions to warrant any shattering "policy relevant findings." While the study provides interesting data, it still leaves me as a school superintendent with a great many questions about the sources of racial tension in the classroom and about possible measures to reduce tension.

In her paper, Eleanor Leacock devotes a substantial amount of attention to discussing what she terms the "we-they" dichotomy. I believe there is considerable danger in accepting such a dichotomy as the pervasive thought process underlying modes of behavior. It does not leave much room for assessing the fall-out between the extremes of the dichotomy and, I fear, limits the objectivity of researchers and educators in dealing with the basic issue of race relations in the schools. Indeed, one can even challenge Arensberg's position, as cited by Leacock, that Americans make "two-fold judgments based on principle": moral-immoral, legal-illegal, right-wrong, etc. By contrast, one could make a strong case that Americans tend to judge people on a continuum and not necessarily at the extremes of a dichotomy.

Much of the paper appears to me to be an attempt to sell a certain philosophy. While interesting, it left me with questions about the implications for research methodology. Leacock's hypothesis that "true valuing of cultural differences is inseparable from true valuing of individual differences" should be tested. Can it best be tested by quantitative or qualitative methodology? And while I agree that this valuing "is necessary if the educational principles advocated in teacher training are ever to be applied," what qualitative and/or quantitative research supports this statement? Valuing cultural pluralism may be most helpful, but it does not guarantee any change that will help minority and poor children learn at a level of which they are capable.

Leacock's suggestions for further research also leave me with questions about methodology. She suggests several interesting and researchable issues and recommends generally a need "to analyze and document both divisive and undermining techniques unwittingly used by teachers...and the styles of teachers who are successful in heterogeneous and minority classrooms." Citing her own research in evidence, Leacock seems to suggest that "key incidents and teacher statements" can be used to summarize and characterize such differences among classrooms and teaching styles. But who judges the key incidents and teacher statements? How does one then generalize from such situations in order to develop alternative models for teachers? It would seem to me that such studies would require considerable qualitative exploration, supplemented with some rather careful quantitative research. However, this question of methodological approach is not pursued in the paper.

Rather than identifying or describing a methodology for assessing race relations, the Leacock paper presents the author's theory of race relations as arising from a dichotomous value framework. It is an attempt to explain why there are problems in race relations–not a discussion of how we might investigate the problems. The paper does, however, raise several intriguing questions that should be explored and that might produce useful insights that could be translated into better decisions about what happens in the classroom.

WORKSHOP ATTENDEES

Blandina Cardenas
Intercultural Research Assn.
San Antonio, Texas

Courtney B. Cazden
Harvard University

John H. Chilcott
University of Arizona

William W. Cooley
University of Pittsburgh

James Cooper
University of Houston

Robert L. Crain
The Rand Corporation
Santa Monica, California

John E. Dawson
Naval Postgraduate School

Mary B. Dawson
Far West Laboratory
for Educational Research
and Development

Elliot W. Eisner
Stanford University

Frederick Erickson
Harvard University

George W. Fairgrieve
The Meramec School
Clayton, Missouri

Stephen E. Fienberg
University of Minnesota
St. Paul, Minnesota

Kenneth Fishell
University of Vermont

Gerald Fry
Stanford University

Gary A. Griffin
Teachers College
Columbia University

Gene E. Hall
University of Texas

Gladys G. Handy
Pennsylvania Dept. Of
Education

John K. Hemphill
Far West Laboratory
for Educational Research
and Development

Michael Hendricks
Northwestern University

Robert Herriott
ABT Associates

John D. Herzog
Northeastern University

Paul D. Hood
Far West Laboratory
for Educational Research
and Development

Dell H. Hymes
University of Pennsylvania

Rex Jackson
Educational Testing Service

Lynn Jenks
Far West Laboratory
for Educational Research
and Development

Virginia Koehler
National Institute of Education

Judith Lanier
Michigan State University

Andrea Lash
Far West Laboratory
for Educational Research
and Development

Eleanor B. Leacock
The City College of The
City University of New York

Marion B. Lentz
Far West Laboratory
for Educational Research
and Development

Robert C. McClure
National Education Association

Bernard McKenna
National Education Association

Mary Ann Millsap
National Institute of Education

Frederick Mulhauser
National Institute of Education

Paul Nachtigal
Consultant to the
Ford Foundation

Melvin R. Novick
University of Iowa

Andrew Porter
National Institute of Education

Ray C. Rist
National Institute of Education

Judith Ruchkin
University of Maryland

Nancy Safer
Bureau of Education
for the Handicapped

John Schwille
National Institute of Education

Lee Shulman
Michigan State University

Roger W. Shuy
Georgetown University

Louis M. Smith
Washington University
St. Louis, Missouri

Roulette W. Smith
Far West Laboratory
for Educational Research
and Development

Lance R. Sobel
University of Washington

Frank Sobol
National Institute of Education

Jane Stallings
Stanford Reasearch Institute

Walter J. Symons
Alum Rock Union Elementary
School District

Robert B. Textor
Stanford University

William J. Tikunoff
Far West Laboratory
for Educational Research
and Development

Tom Tomlinson
National Institute of Education

Kent Viehoever
National Institute of Education

Beatrice A. Ward
Far West Laboratory
for Educational Research
and Development

Robert E. Wentz
Saint Louis Public Schools

COUNCIL ON ANTHROPOLOGY AND EDUCATION
Board of Directors 1976-77

President: *Frederick D. Erickson*, Harvard Graduate School of Education, Longfellow Hall, 13 Appian Way, Cambridge, MA 02138

1st Vice President (Immediate Past President): *John D. Herzog*, Northeastern University, College of Education, Boston, MA 02115

2nd Vice President (President Elect): *Dell Hymes*, Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19174

Secretary-Treasuer: *Glenn L. Hendricks*, Student Life Studies and Planning, 328 Walter Library, University of Minnesota, 117 Pleasant Street S.E., Minneapolis, MN 55455

Editor, Anthropology and Education Quarterly: *Charles Harrington*, Box 19, Teachers College, Columbia University, New York, NY 10027

Members at Large: *Peggy Sanday* (1977), Department of Anthropology, University of Pennsylvania, 33rd and Spruce Streets, Philadelphia, PA 19174; *Courtney Cazden* (1978), Graduate School of Education, Larsen Hall, Harvard University, Cambridge, MA 02138; *St. Clair Drake* (1979), Department of Anthropology, Stanford University, Sanford, CA 94305

Standing Committee Chairpersons

Committee 1: Anthropological Studies of School and Culture—
Kenneth Baer, The Lullwater School 2399 Williams Lane, Decatur, GA 30033
Richard Zimmer, Hutchin School at Sonoma State, Rohnert Park, CA 94928

Committee 2: Cognitive and Linguistic Studies—
John Ogbu, Department of Anthropology, University of California, Berkeley, CA 94820
Evelyn Jacob, Department of Anthropology, University of Pennsylvania, Philadelphia, PA 19174

Committee 3: Materials and Methods for Teaching Anthropology—
Marion J. Rice, College of Education, University of Georgia, Athens, GA 30602
Thomas Dynneson, University of Texas, Permian Basin Parkway, E. University, Odessa, TX 79762

Committee 4: Ethnographic Approaches to Evaluation in Education—
Woody Clark, Institute for Race and Community Relations, 30 Northgate Hall, University of California, Berkeley, CA 94720
Jean J. Schensul, Kings Point, Monaco N668, Del Ray Beach, FL 33446

Committee 5: Transnational Issues in Education and Change—
Johannes Wilbert, Department of Anthropology, University of California, Los Angeles, CA 90024

Committee 7: Blacks in Education—
Sheila Walker, Division of Higher Education, 4607 Tolman Hall, University of California, Berkeley, CA 94720

Committee 8: Spanish-Speaking Concerns in Education—
Lucila Carrasco-Schoch, Sociology Department, University of California at San Diego, La Jolla, CA 92037

Committee 9: Women in Schools and Society—
Judith Preissle Goetz, Social Science Education, Dudley Hall, University of Georgia, Athens, GA 30602

Committee 10: Environmental Issues in Anthropology—
Kathleen Adams, Department of Anthropology, Central Washington State College, Ellensburg, WA 98926

Ad Hoc Committee Chairperson

Ad Hoc Committee 1: Employment
Willis Sibley, Department of Anthropology, Cleveland State University, Cleveland, OH 44115

COUNCIL ON ANTHROPOLOGY AND EDUCATION MEMBERSHIP/SUBSCRIPTION FORM

Name ______________________

Address ______________________

____

____
city state zip code

Indicate Committees of Special Interest:

___ Individual Membership $7.50 (___ AAA Member) includes $4.00 for a subscription to the *Quarterly*

___ Library Subscription $10.00

Mail with check to Council on Anthropology and Education, 1703 New Hampshire Ave., N.W., Washington, DC 20009

ANTHROPOLOGY AND EDUCATION QUARTERLY
Council on Anthropology and Education
1703 New Hampshire Avenue NW
Washington DC 20009

2nd Class
Postage Paid
Washington DC

CA 2
ERIC CHESS
DIR
855 BROADWAY
BOULDER, CO 80302

1299 530780